Does PPP hold for Norway?

Einar Wøien Nordbø[1]

Master thesis for the cand.oecon.-degree

University of Oslo, Department of Economics

16th December 2004

[1] *Contact information:* Work: Forskningsavdelingen, Norges Bank, PB 1179 Sentrum, 0107 Oslo. Phone: 22316295 E-mail: einar.nordbo@norges-bank.no.

Contents

List of Figures

List of Tables

Abstract

This paper is an empirical investigation of whether the theory of purchasing power parity (PPP) describes Norwegian data well. I have used absolute price data from the Penn World Table (PWT) and constructed real effective exchange rates (REER) for Norway against various groups of countries. My main focus is on an importweighted REER against a group of 40 countries. By employing simple unit root tests on this REER over the period from 1973 to 2000 I get strong rejections of the unit root hypothesis, and accordingly, firm support of PPP. The implied half-life of deviations from PPP is less than two years. This is contrary to what some previous PPP studies done on Norwegian data have found. Furthermore, I test for PPP also bilaterally, but I can only detect evidence of PPP against a few countries.

The REERs thus appear to be more stationary than most bilateral real exchange rates. This feature can be explained if it is the case for Norway that shocks to some bilateral real exchange rates are positive, whereas others simultaneously are negative. In a weighted average of several bilateral real exchange rate - a REER - these shocks may roughly cancel, making the REER more stationary.

Possible explanations for why Norway has exhibited faster mean reversion than many other countries is the policy of nominal exchange rate stability that hase been followed in the post Bretton Woods period, and the coordinated policy attempts at maintaining the competitiveness of the tradable sector that has been done.

In addition, I have tried to identify factors that explain the variability in half-life of deviations from PPP as compared with different countries. I have found that distance, the level of development and the level of bilateral trade are possible explanatory factors. Looking at nearly two centuries of data I find the Norwegian-UK real exchange rate to be trend stationary. This can be interpreted as evidence of a Balassa-Samuelson variant of PPP, as productivity growth - here measured by growth in real gross domestic product (GDP) per capita - on average has been higher in Norway than in the UK in the respective period.

Furthermore, I have tested for nonlinearities in some bilateral real exchange rates for Norway, as nonlinear models have been found to describe some other real exchange rates better than linear models. But the tests I have performed do not indicate any meaningful nonlinearity.

Abbreviations and acronyms

Abbreviations and acronyms frequently used in this paper are explained here. Most of them are explained the first time they appear in the paper as well, but I have still included this list to make it easier for the reader.

ADF Augmented Dickey-Fuller, a frequently used test for non-stationarity.

ADF-coefficient The level variable coefficient in an ADF regression. The parameter ϱ in equation 2.6.

CW(14) A real effective exchange rate constructed by the competitiveness weights of the OECD. Constructed for comparison. Countries listed in the appendix, section A.4.

CW24 A real effective exchange rate constructed by the competitiveness weights of the OECD. Included countries listed in the appendix, A.4.

DF-GLS Dickey-Fuller, generalized least squares. Another test for non-stationarity.

ESTAR Exponential smooth transition autoregressive, a nonlinear model.

I40 The broadest real effective exchange rate index used in this paper, constructed by import weights. Includes 40 countries. The countries are listed in the appendix, A.4

I44 A nominal effective exchange rate constructed by Norges Bank, using import weights. See the appendix, A.4.

IFS International Financial Statistics, a database published by the IMF.

IMF International Monetary Fund

IW(14) An importweighted real effective exchange rate constructed in this paper for comparison purposes. See appendix A.4.

KPSS Kwiatkowski, Phillips, Schmidt and Shin. A test of stationarity.

LSTAR Logistic smooth transition autoregressive, a nonlinear model.

NID Normally, independently distributed.

OECD The Organization for Economic Co-operation and Development.

PP Phillips-Perron, a test for non-stationarity.

PWT Penn World Table.

REER Real effective exchange rate.

s.e. Standard error. The standard errors of the estimated parameters are normally reported in (parentheses) in the tables, to distinguish them from p-values, which are reported in [square brackets].

STAR Smooth transition autoregressive, a family of nonlinear models. ESTAR and LSTAR are two members.

TAR Threshold autoregressive, a nonlinear model.

TW(14) A tradeweighted real effective exchange rate constructed for comparison. The included countries are listed in the appendix, A.4.

TW16 A tradeweighted real effective exchange rate. The included countries are listed in the appendix, A.4.

****** Denotes significance at the 1% level. A single * and (*) implies significance at the 5% and 10% level, respectively, for the results reported in the tables.

hl(ADF) Half life calculated by the simple formula, see section A.2 in the appendix, on the ADF-coefficient. hl(IR) and hl(AR1) denote half-lives based on the impulse response function (IR) and an estimated AR(1) model.

PPP Purchasing power parity.

GDP Gross domestic product.

AIC Akaike information criterion.

MAIC Modified Akaike information criterion.

BIC Bayesian information criterion.

Preface

This paper was written while I had a student internship in the Research Department at Norges Bank[1] (the central bank of Norway) from June 2004 through November the same year. I would like to express my gratitude to Norges Bank for providing me with economic funding and inspiring working conditions during this period, and above all, for allowing me to write my master thesis about a topic I found interesting. I would in particular like to thank my advisor in Norges Bank, Farooq Akram, who helped me with this paper from the very beginning and gave me a lot of good advice throughout the process. But I am also indebted to Øyvind Eitrheim, Ida Wolden Bache and Lucio Sarno for meaningful discussions and hints along the way. Furthermore, I am grateful to Katrine Godding Boye for allowing me to use some of the data she had gathered. I would also like to thank Maurice Obstfeld and Michael Jansson, both at UC Berkeley, for giving me some advice on relevant literature. Their classes in international economics and time series analysis, which I followed while I was a visiting student at Berkeley from August 2003 till June 2004, were also the classes that I have benefitted most from when writing this thesis. And finally I must thank my advisor at the University of Oslo, professor Erik Biørn, who read numerous drafts of this paper and helped me detect many mistakes. Without the help and advice from all these people, this paper would definitely have been much worse than it actually is. But all remaining errors are of course my own responsibility.

[1]The views expressed in this paper are those of the author and not necessarily those of Norges Bank.

Chapter 1

Introduction

Under the skin of any international economist lies a deep-seated belief in some variant of the PPP theory of the exchange rate.

Dornbusch and Krugman (1976)

The topic of this paper is purchasing power parity - the simple idea that the price levels in different countries should be equal when they are converted to a common currency. The literature on this issue has seen an explosive growth in recent decades, and as argued by Imbs et al. (2005, forthcoming), PPP is perhaps the most intensely researched area in international macroeconomics.

One reason for this is probably that researchers for a long period failed to detect any evidence of PPP at all, and they were thus contradicting a basic assumption in many standard models. As new econometric tools were developed and researchers started using longer time series of data, most studies would be more supportive of PPP, but only as a long run property of the real exchange rate. A consensus was built around an estimated half-life of deviations from PPP on 3 to 5 years. This remarkable persistence of shocks to the real exchange rate combined with the very high short run volatility, led Rogoff (1996) to formulate "the purchasing power parity puzzle". The basic rationale for this puzzle is that the short run volatility is so large that it can only be accounted for by monetary shocks. But if monetary factors were the main source of shocks to the real exchange rate, one would expect that the shocks would die out much faster, as monetary shocks only affects the real economy through nominal rigidities in standard models.

There have of course been numerous attempts at explaining this puzzle, and some of them are described in the following chapter. The purpose of this paper is rather to investigate the PPP issue with a basis in Norwegian data. This has been done before, for instance by Bjørnland and Hungnes (2003) and Akram (2005, forthcoming), but no concensus has been reached. This paper differs from the two mentioned above because it uses absolute price data from the Penn World Table, which makes it possible to test also the absolute version of PPP. The Penn World Table is the best attempt so far at determining the price of the same basket of goods across countries, denoted in a common currency.

I have studied PPP by looking at real effective exchange rates against various groups of countries over the post Bretton Woods period, and the broadest group, named I40 here, consists of 40 countries, all included in the I44-index constructed by Norges Bank. But I have also studied PPP bilaterally against the same 40 countries and given an estimate of the speed of convergence to PPP for each country. Furthermore, I have tried to identify factors that can explain the variability in the speed of convergence to PPP. This gives a broader basis for discussing PPP with respect to Norway than in previous studies.

But I have also looked at longer time series, and the longest one is the real exhange rate between Norway and the UK from 1819 to 2003. In addition, I have tested for nonlinear mean reversion in some bilateral real exchange rates, one of the possible solutions to the so-called PPP puzzle that has been put forward.

The majority of the published PPP studies have focused on the relationship between the major international currencies. Norway differs from these economies in several ways. Above all, Norway is a small open economy and is therefore much more influenced by developments abroad than for instance the US. In addition, although Norway is a highly developed country by most standards, its export has a much larger share of non-manufactured goods (mainly oil) than many other developed economies. A related factor is that Norway, through the oil discoveries in the 1970s and the resulting revenue, has been able to finance an extensive public spending. These are some factors that can cause PPP to be a more or less relevant concept for Norway than for other countries - and thus, they provide some reasons for why a PPP study with a basis in Norwegian data can be interesting.

This paper is organized in the following manner. Chapter 2 describes the PPP theory and surveys the empirical literature. I also provide some theoretical considerations for why PPP might fail. In chapter 3 I report the data I have used, how they have been aggregated and the econometric methods I have employed to test for PPP. The 4th chapter contains the empirical results of my studies. Chapter 5 concludes and summarizes the paper.

The econometric software programs I have used in this paper are Eviews 5.0 and PcGive 10.

Chapter 2

Theory and empirical literature on PPP

2.1 Economic Theory and Concepts

The theory of purchasing power parity is based on the assumption of frictionless goods market arbitrage. When the price of a similar good in two different countries is converted to a common currency, the price should be equal. This is often referred to as the *law of one price*. In mathematical terms, this can be expressed in the following formula for a good i

$$P_i = EP_i^*, \tag{2.1}$$

where P_i is the domestic price of the good, E is the exchange rate, and P_i^* is the foreign price. If this condition is violated, economic reasoning suggests that it would be profitable to export the good from the country where it is cheap to where it is expensive. And if prices on a given day differed by a substantial amount, for instance due to a recent devaluation, one would expect them to be equalized by arbitrage as traders take advantage of the difference.

Of course, one may point to a number of reasons for why this condition might fail. Two of the most obvious are transportation costs and tariff barriers. But the idea that changes in the real exchange rate, given by $R = EP^*/P$, where P denotes the aggregate price level, should be completely unpredictable, as suggested by Roll (1979), is hard to accept for most economists. Roll argued that the real exchange rate should behave as an asset price if foreign exchange markets are efficient. But as pointed out by Froot and Rogoff (1995), the real exchange rate is not a price of a traded asset. The asset market analogy is therefore somewhat faulted.

If the law of one price holds for all commodities it implies that *absolute purchasing power parity* is valid as well. Absolute PPP requires that a given basket of goods and services costs the same in two countries

$$\sum_{i=1}^{n} \alpha_i P_i = E \sum_{i=1}^{n} \alpha_i P_i^*, \ \sum_{i=1}^{n} \alpha_i = 1. \tag{2.2}$$

α_i is the weight assigned to the different items in the index. As can be seen, absolute PPP may hold even though individual goods prices differ.

Testing this condition on actual data has proved to be difficult. One problem is lack of comparable data. The construction of a consumer price index is usually a task done by the national statistical authorities, and the method and weights used are therefore countryspecific. An additional problem is that consumer price indices are just what they are called, an index relative to a base year. Unless absolute PPP prevailed in the base year, consumer price indices don't inform us about the magnitude of absolute price deviations from one country to another.

The Penn World Table (Heston, Summers and Aten, 2002) used in this paper is the best attempt so far to circumvent this problem. It contains data on the price in US dollars of a comparable basket of goods for more than 160 countries, stretching back to 1950 for most of them.[1] But the PWT has not been used extensively in tests of PPP. The huge majority of PPP studies have instead concentrated on so-called *relative PPP*. This condition is satisfied if the relationship between the foreign and domestic price level, measured in a common currency, is constant over time. It is usually not required that the weights in the two indices are equal.

$$\sum_{i=1}^{n} \alpha_i P_{it} / \sum_{i=1}^{n} \alpha_i P_{it-1} = (E_t/E_{t-1}) (\sum_{i=1}^{n^*} \alpha_i^* P_{it}^* / \sum_{i=1}^{n^*} \alpha_i^* P_{it-1}^*). \tag{2.3}$$

The relative version of PPP allows for transportation costs and other obstacles to trade, since the real exchange rate is not required to be equal to one. But this version of PPP cannot be derived directly from the law of one price (Dornbusch, 1985). An alternative

[1]The Penn Word Table is described in more detail in section A.1 in the appendix.

theoretical underpinning is the neutrality property of money in the long run, which says that a purely monetary disturbance eventually will result in an equal change in the money stock and all prices, including the exchange rate. The real exchange rate will therefore be unchanged. The belief in PPP is in this respect closely related to the belief that the majority of shocks to the real exchange rate are monetary disturbances.

2.2 The history of PPP

The idea of PPP is not new. The origins of the theory can be traced as far back as to the Salamanca school in 16th century Spain, and classical economists like David Ricardo, John Stuart Mill and Alfred Marshall have all discussed variants of PPP. But the modern version of the theory is attributed to the Swedish economist Gustav Cassel, who was the one who came up with the name "Purchasing Power Parity". He also became a leading protagonist of PPP as a theory of exchange rate determination in the period after World War I, when economists and politicians discussed if and how one should return to the gold standard. Most countries had allowed their currencies to float during the war, and the inflation rates had been widely different from country to country. Cassel proposed using PPP-calculations to find the appropriate level to reinstate gold parities, instead of restoring the prewar parities, as others argued. The influential John Manyard Keynes was among the economists who gave him support, but Cassel's ideas remained controversial.

PPP reemerged as a relevant theory in the aftermath of World War II, when once again exchange rates had to be reset following a period of no convertibility. An important theoretical innovation came with the work of Balassa (1964) and Samuelson (1964), who predicted that countries that had a relatively more rapid productivity growth in the tradable sector would experience a real appreciation of the currency.[2]

The monetary theory of the exchange rate, which was popular in the early 1970s,[3] had as a key assumption that purchasing power parity held at all times. But the downfall of the Bretton Woods system in 1973 led to an unprecedented degree of real exchange rate volatility, and the deviations from PPP were both large and persistent. A number of empirical studies

[2]See section 2.4 for a more detailed description of the Balassa-Samuelson effect.

[3]See Frenkel and Johnson (1976) for more on the monetary theory of the exchange rate.

were even unable to reject the hypothesis of the real exchange rate as a random walk. More recent work have been more supportive of PPP as a long-run property of the real exchange rate, and as Taylor and Taylor (2004) note in their recent survey of the PPP literature:

> *In this respect, the idea of long-run PPP now enjoys perhaps its strongest support in more than thirty years, a distinct reversion in economic thought.*

The empirical work on PPP will be explored more closely in the following section.

2.3 Empirical work on PPP

2.3.1 The law of one price

Although the idea of purchasing power parity sounds plausible, the most striking aspect of the empirical literature on PPP is probably how large and persistent the deviations can actually be. A famous example of deviations from the law of one prize is the "Big Mac index" published by the magazine "The Economist", see table 2.1. A Big Mac hamburger in different countries is very close to being a homogenous good, but in spite of this, the most expensive burger in the sample, the one you find in Iceland, costs almost five times as much as the cheapest one, which is sold in the Philippines. But the large deviations are not that difficult to understand once you consider the huge nontraded components of the price of a Big Mac, like ingredients bought locally, wages to the salespeople and property rental costs.

In an entertaining paper Cumby (1996) finds that once you correct for average deviations, that is, use the "Big Mac index" as a measure of relative PPP, the index is actually a pretty good indicator of future exchange rate changes. Half-life[4] of deviations from "Big Mac-parity" is found to be about one year, far below the consensus half-life estimate of 3-5 years in the literature.

But the law of one price applies to some markets. Figure 2.1 shows the market price differential for one ounce fine gold in New York and London, measured in US dollars, in percentage of the London price for every year from 1791 to 1998. The biggest price difference is 13 percent, recorded just after World War II.[5] The average deviation, in absolute value,

[4] The time it takes before the effect of a unit innovation is halved, see section A.2 in the appendix

[5] The data are gathered from Officer (2001, 2002).

Table 2.1: The Big Mac Index

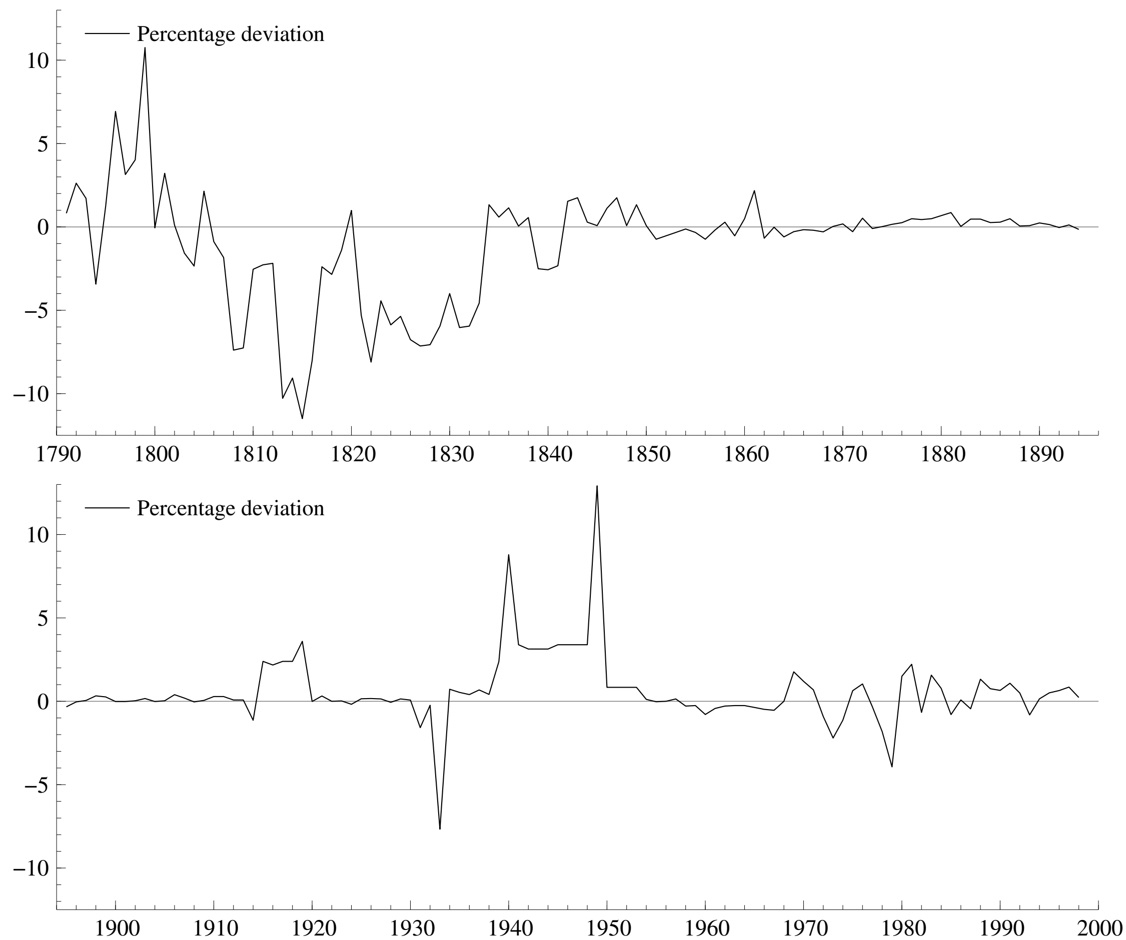
Country	*Big Mac price*
United States	2.90
Argentina	1.48
Australia	2.27
Brazil	1.70
UK	3.37
Canada	2.33
Chile	2.18
China	1.26
Czech Rep.	2.13
Denmark	4.46
Euro area	3.28
Hong Kong	1.54
Hungary	2.52
Iceland	6.01
Indonesia	1.77
Japan	2.33
Malaysia	1.33
Mexico	2.08
Norway	5.18
Philippines	1.23
Poland	1.63
Russia	1.45
Singapore	1.92
South Africa	1.86
South Korea	2.72
Sweden	3.94
Switzerland	4.90
Taiwan	2.24
Thailand	1.45
Turkey	2.58
Venezuela	1.48

Prices of a Big Mac in current US dollars. Source: Economist.com, May 27th 2004.

is 1,6 percent, and the median deviation 0,6 percent. The average deviation is somewhat higher before 1900 compared to after 1900, 2,0 percent versus 1,2 percent.

In a frequently cited study on the law of one price, Engel and Rogers (1996) investigate relative price differences between US and Canadian cities. They find that both whether the cities are located on different sides of the border and the distance between them help explain the deviations. But just crossing the border in itself has a substantial impact on the

Figure 2.1: The gold price differential US - UK



The difference in the gold price in current US dollars in New York and London, in percentage of the London price. I have annual data from the period 1791 to 1998.

volatility of the price differential between two cities. It is equivalent to a distance of 75 000 miles, according to one of their calculations.

In another study of the law of one price, Froot, Kim and Rogoff (1995) challenge the conventional belief that the high volatility of relative prices is a characteristic of the modern period of floating exchange rates only. Looking at annual commodity price data spanning seven centuries for England and Holland, they find that the volatility and the persistence of deviations from the law of one price have been relatively stable over time.

These are only two of a large number of studies that question the applicability of the law of one price. Given the weak empirical support for the basic building block of the PPP theory, it should not come as a surprise that researchers have had problems with producing strong evidence in favour of PPP also on the aggregate level.

2.3.2 PPP on the aggregate level

I will divide the empirical literature on PPP into four groups. More extensive surveys are given in Froot and Rogoff (1995), Rogoff (1996), Sarno and Taylor (2002) and Taylor and Taylor (2004). This survey is inspired by all the surveys mentioned, but draws most heavily on Froot and Rogoff (1995).

In what I will refer to as the first group in the empirical PPP literature, consisting of the early tests, the null hypothesis is that PPP holds. One early study that was supportive of PPP, was done by Frenkel (1978). He ran the following regression,[6]

$$e_t = \alpha + \beta(p_t - p_t^*) + \varepsilon_t, \tag{2.4}$$

on data from several hyperinflationary economies, and he found estimates of the coefficient β close to one, which he interpreted in favour of PPP.[7] But most of the other early studies were not supportive of PPP. A general problem with these studies was the lack of adequate econometric techniques. Today we know that if the error term in (2.4) is integrated of order one, I(1),[8] the standard OLS test of $\beta = 1$ is invalid.[9]

The second group of studies, where the null hypothesis is that the real exchange rate is a random walk, takes advantage of the work by Dickey and Fuller (1979) and others in developing a valid test procedure for regressions containing a unit root. One example of a

[6]Throughout, lowercase letters denote natural logs.

[7]If β is equal to one, it implies that the real exchange rate will be equal to constant, which is equivalent with the relative version of PPP. Absolute PPP would prevail if the constant equaled zero, given that you have absolute price data and the same basket of goods.

[8]As it would be if the autoregressive operator $\psi(L)$ in the following expression,

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)\varepsilon_t \equiv \psi(L)\varepsilon_t = u_t, \tag{2.5}$$

has a unit root, where u_t is white noise. This implies that the series does not have constant mean and variance.

[9]For a good introduction to regressions containing a unit root, see Hamilton (1994), chapter 17. A more advanced treatment is given in Stock (1994).

paper where this approach is used is Meese and Rogoff (1988). The authors cannot reject the null hypothesis of the real exchange rate being integrated of order one for dollar-mark, dollar-sterling and dollar-yen series over the post Bretton Woods period. Meese and Rogoff use a so-called augmented Dickey-Fuller (ADF) test, with the following specification

$$\Delta r_t = \mu + (\varrho - 1)r_{t-1} + \sum_{i=1}^{k} \psi_i \Delta r_{t-i} + \varepsilon_t, \tag{2.6}$$

where r_t is the log of the real exchange rate.

Given the low power of the unit root test, it is not surprising that Meese and Rogoff and others who employed similar tests failed to find evidence of PPP.[10] Alternative tests using variance ratios and tests of fractional integration were performed with the same depressing results for the PPP advocates.

The profession has followed two different paths in trying to resolve the problem of low power. One solution has been to extend the sample period. Frankel (1986) used data on the dollar-sterling real exchange rate from 1869 to 1984, and was able to reject a random walk. His half-life estimate was 4,6 years. Johnson (1990) also found support of PPP using cointegration methods on 120 years of US/Canadian real exchange rate data. Edison and Klovland (1987) looked at annual data from Norway and the UK spanning the period from 1874 to 1971 and detected evidence in favour of long-run PPP if they allowed for different short run dynamics during the floating period from 1914 to 1928.

An obvious problem with these studies is that they span periods with varying exchange rate regimes. It is well documented that the real exchange rate variability tends to be higher during periods with a floating exchange rate (see Mussa, 1986). Lothian and Taylor (1996) investigate the implications this has for studies using long data sets. Using sterling-dollar and sterling-franc data spanning two centuries, they cannot reject the hypothesis of no structural change before and after Bretton Woods. This indicates that running regressions on data from periods with different exchange rate regimes may be a valid procedure, but it is nevertheless a potential source of misleading conclusions. Another concern with these studies was raised

[10] In section 4.2 I assume that the true process is AR(1) and use the asymptotic variance of the dependent variable to calculate how many years of data which are necessary to reject the null hypothesis with a standard Dickey-Fuller t-test. If the true half-life is one year, 25 years of data are necessary, for instance.

by Froot and Rogoff (1995). Since the countries where the long run data series are easiest available, tend to have been continuously among the world's most developed, these studies might exaggerate the evidence of long run PPP. This concern is particularly relevant if the Balassa-Samuelson effect is true.

The second approach to solve the problem of low power has been to extend the number of observations by using data from more than one country for the post Bretton Woods period. But many papers within this strand of the literature suffer from a serious weakness, as pointed out by Sarno and Taylor (2002). If a unit root test using panel data results in a rejection of the null hypothesis of a random walk, the only valid conclusion is that at least one of the series used is stationary. One cannot infer from this that PPP holds in general.

One example of a study employing this procedure is Abuaf and Jorion (1990). They use real exchange rate data for ten countries versus the US over the period from 1973 to 1987, and are able to reject the null hypothesis of joint nonstationarity, but only at a significance level of ten percent. Their estimated half-life of deviations is between 3 and 5 years.

Most of the studies in the second group employed univariate test statistics to investigate the PPP issue. In what I will refer to as the third group of the literature on PPP, the more modern method of cointegration is used, taking advantage of the work of Engle and Granger (1987).

The basic idea that warrants the use of cointegration techniques in PPP studies is the belief that the exchange rate e_t and the price differential $p_t^* - p_t$ both are I(1), but that a linear combination of them is stationary. If this is the case, we say that the series are cointegrated. The null hypothesis in these types of studies is therefore that no linear combination of the exchange rate, the foreign and the domestic price level is stationary

$$e_t - \beta p_t + \beta^* p_t^*. \tag{2.7}$$

The only difference from the simple unit root tests is that I there imposed the restrictions $\beta = \beta^* = 1$ beforehand and checked for stationarity of the real exchange rate

$$r_t = e_t - p_t + p_t^*. \tag{2.8}$$

An important principle in econometric testing is that one should test a theory, not just run random regressions. By allowing the coefficients in (2.7) to be different from one, the researcher should have a clear idea of why they might differ. If the series turn out to be cointegrated, but the coefficient estimates differ clearly from one, without any plausible economic explanation, the researcher cannot really claim to have found evidence of PPP.

Some theoretical considerations for why the coefficients might be different from one is provided by Taylor (1988). He constructs models with transportation costs and measurement error and shows that this might give coefficients different from one. His study is among the first to apply cointegration methods to the PPP theory, but he is unable to reject the null hypothesis of no cointegration for any of the country pairs he considers.

Later studies using cointegration methods have been more supportive of PPP, but the estimates of β and β^* vary wildly and it is very hard to provide any theoretical explanation for why this should be the case. One potential source of the strange estimates is small sample bias, which has been known as a potentially substantial problem in cointegration studies since the work of Banerjee et al. (1986).

In what I will refer to as the fourth group of PPP test, nonlinear estimation techniques are applied. This is because economic reasoning suggests that real exchange rate dynamics may be described better with a nonlinear model. Transportation costs, tariffs and other barriers to trade have been mentioned before as possible explanations for deviations from the law of one price. Although prices measured in a common currency differ, the marginal profit of shipping goods between the countries might be less than the cost. In this respect the relative prices in the two countries can be disconnected. But if the price deviations become large enough, it will be profitable to ship goods, and the price differences should be reduced. Obstfeld and Taylor (1997) find evidence of this effect using a threshold autoregressive model (TAR) on disaggregated price data. The idea underlying the use of this particular model is that relative prices may be nonstationary within a band, but that it becomes a mean reverting process once it crosses one of the thresholds.

Other reasons for why the adjustment process might be nonlinear have also been presented. Kilian and Taylor (2003) suggest that heterogeneous opinions in the foreign exchange market about the equilibrium level of the nominal exchange rate might create nonlinearities, and Taylor (2004) argues that the intervention operations of central banks can have the same effect. The literature on nonlinearities is explored further in section 4.5.

2.4 Should PPP hold?

The large number of studies that documented substantial deviations from PPP and only provided empirical support of PPP in the very long run, have led economists to come up with various suggestions for why this might be the case. Before I start examining the different theories trying to explain the deviations from PPP, it might be useful to rewrite the basic PPP equation in order to understand where the deviations might come from. The expression for the real exchange rate in natural logs was given in (2.8). Both p^* and p are national price indices consisting of both tradable and nontradable goods. Assuming that the weights are equal in both countries (although differing weights may also be a source of PPP deviations), I obtain the following, where subscripts N and T denote nontradable and tradable goods, respectively

$$p = \alpha p_T + (1 - \alpha)p_N \tag{2.9}$$

and

$$p^* = \alpha p_T^* + (1 - \alpha)p_N^*. \tag{2.10}$$

Inserting these two expressions into equation (2.8), I obtain the following expression for the real exchange rate. The time subscripts have been dropped for simplicity

$$r = \alpha(e + p_T^* - p_T) + (1 - \alpha)(e + p_N^* - p_N). \tag{2.11}$$

Since the real exchange rate here is on log form, absolute PPP requires that $r = 0$, whereas relative PPP holds if the real exchange rate is equal to a constant. If I make the unrealistic assumption that the law of one price holds for tradables, implying that $(e + p_T^* - p_T) = 0$, we notice that we may still have deviations from PPP if the price of nontradables differ. The matter is further complicated by the fact that most goods that are highly tradable, will have a significant nontradable component.

The most famous theory explaining why the price of nontradables might differ, is the

Balassa-Samuelson effect. I will not derive the theory formally here,[11] but only explain its main features in words.

The model underlying the theory assumes that you are looking at a small open economy producing only two goods, one tradable and one nontradable. The law of one price is assumed to hold for tradables, and also the nontradable goods markets are perfectly competitive. The production function exhibits constant returns to scale in the two input factors, capital and labour. Capital is mobile between sectors and internationally, which implies that the real interest rate is determined on the world market. Labour is mobile between sectors, but not across borders. This means that the wage will be the same in both sectors, but can vary internationally.

The key prediction of the model is that a country will experience a real appreciation compared with another country if its productivity growth advantage in the tradable sector exceeds its productivity growth advantage in the nontradable sector. This is conditional on the plausible assumption that the nontradable sector is labour intensive, i.e., labour's share of the income generated is at least as high in the nontradable as in the tradable sector. That the productivity growth advantage is relatively higher in the tradable sector is reasonable given that the nontradable sector contains most of the service industries (e.g. haircuts). And if the nontradable sector is labour intensive, you may see a real appreciation even if the productivity growth advantage in the two sectors are balanced.

To provide some economic intuition for the model, let us start in the tradable sector. For simplicity I assume that the exchange rate is fixed, and that all economic variables in the two countries initially are identical. Since the law of one price holds, the price of the tradable good will continuously be the same in both countries. Without loss of generality, I can assume that the tradable price stays fixed. In the country where the productivity growth in the tradable sector is higher, a larger part of the tradable price will then go to the workers as wages. Since the wage level by assumption must be the same in both sectors, this means that the country with the productivity growth advantage in the tradable sector will pay its workers in the nontradable sector relatively better. And since the productivity growth in the nontradable sector has been smaller, prices here must increase, which results

[11] For a formal derivation, see for instance Obstfeld and Rogoff (1996), pp. 204–214.

in a general price increase and a real appreciation of the currency.

The empirical support for the Balassa-Samuelson effect is somewhat mixed, though. The theory works very well for Japan, but studies done on other countries draw a less clear-cut picture. Since the countries that have enjoyed a relative productivity growth advantage in the tradable sector will be relatively richer, one way to test for the Balassa-Samuelson effect is to run a regression on the relationship between the price level and income. This is was first done by Balassa (1964) himself for twelve industrial countries in 1960, and he identified a positive relationship that was clearly significant. Other studies have been less supportive. Froot and Rogoff (1991a,1991b) detect only weak correlations, at best, between productivity differentials and the real exchange rate for 22 OECD countries. But Edison and Klovland (1987) are able to find some evidence of a Balassa-Samuelson effect by using long run data for Norway and the United Kingdom. More support is offered by Heston, Nuxoll and Summers (1994), who run a number of regressions using the absolute price data in the Penn World Table. As mentioned before, this is also the data set I am going to use.

The Balassa-Samuelson effect may be the most prominent theory explaining long run deviations from PPP, but numerous other explanations have been put forward. A related theory, which also predicts that richer countries will have a more appreciated real exchange rate, was developed by Kravis and Lipsey (1983) and Bhagwati (1984). Instead of looking at productivity differences, they relax the assumption of perfect international capital mobility. In their model richer countries accumulate more capital per worker, which gives a higher wage level, and a higher price level of nontradables if these goods are labour intensive.

The level of government spending has also been found to explain deviations from PPP. One explanation is that government spending tends to fall more heavily on nontraded goods, and thereby increase their relative price. Both Froot and Rogoff (1991a) and De Gregorio, Giovanni and Wolf (1994) detect government spending as a significant factor in determining the real exchange rate. The effects die out in the long run, but very slowly, with a half-life of more than five years.

Another explanation that is often mentioned is that sustained current account deficits are closely related to long run real exchange rate depreciation. Obstfeld and Rogoff (1995) find some evidence of this. One possible explanation is that[12] current account deficits imply wealth transfers from one country to another, and home and foreign residents may have different consumption patterns.

All of the theories mentioned above seek to explain deviations from PPP assuming that the law of one price holds for tradeables. But as the empirical work has shown, deviations from the law of one price can be both large and persistent. Some of the reasons for this are obvious, like transportation costs and tariff barriers. Even for highly traded goods, a significant amount of the price can be due to nontraded components, as previously discussed. And non-tariff barriers, such as import licensing requirements, technical product standards, anti-dumping duties and so on, may be more important than previously thought.[13] Knetter (1994) do for instance argue that non-tariff barriers is a very relevant factor in explaining high retail prices in Japan.

At a more advanced level, there are two, not mutually excluding, explanations for why the law of one price might fail. "Pricing to market" is one of them. The name was coined by Krugman (1987), and denotes the ability of producers with market power to charge a different price for the same good in various markets. The car industry is often mentioned as an industry where "pricing to market" is particularly relevant. In a world with perfect competition agents could make profits on the price differences across countries, but this is not possible since the producers can separately license the sale of goods in the different countries.

The other approach to explain why a company may charge different prices for the same good in varying countries is offered by Kasa (1992). Instead of focusing on oligopolistic competition, he stresses adjustment costs. It might be costly for the firms to change prices, that is, they face menu costs, or it might be the consumers who face fixed costs in switching between products. In any case, the adjustment cost story provides an alternative to Krugman's rationale for "pricing to market".[14] But as mentioned above, both these theories might be useful to understand what the true process is really like.

[12] Advocated by Krugman (1990).

[13] It is not unlikely that their importance may have grown after international trade liberalization now has reduced traditional obstacles to trade.

[14] But this story can of course not explain long run price deviations.

Chapter 3

Data and method

3.1 Absolute price data

As mentioned earlier, the data set I am going to use in this paper is the Penn World Table. This data set is discussed more closely in the appendix, section A.1. For a more thorough description of the data and how they have been produced, see Summers and Heston (1991). As I also indicated, this data set has not been used extensively in tests of PPP. I am only aware of three studies employing the PWT data - Oh (1996), Parsley and Popper (2001) and Zussman (2002).

One of the reasons for this is probably concern with the quality of the data. As emphasized by Sarno and Taylor (2002), the PWT data are constructed both by interpolations between benchmark years, and extrapolations to countries where no benchmark studies have been done. Because of this they claim that the PWT data become "partially artificial". I find their concern to be somewhat exaggerated. The interpolations between benchmark years are based on national price indices, and it is difficult to see that the difference between using the indices alone and absolute price data based on the indices can be that important. In addition, by using the PWT data, I gain the advantage that I compare prices on the same basket of goods across countries, which is necessary for tests of absolute PPP. The same method has been applied to construct a price index for every country, whereas national statistical authorities use varying methods. By using PWT data I avoid potential problems related to this. Zussman (2002) uses both data from the PWT and data from the International Monetary Fund's (IMF) International Financial Statistics (IFS) database. His results are the same irrespective of which of the two data sources he uses. Another advantage with the PWT is that it has price data for more countries and longer time periods than other data sources. These arguments provide some reasons for why it might be interesting to test PPP by using PWT data.

But more concerns can be raised. One of them is that the PWT only contains annual data. The frequency of the observations does not influence the asymptotic power of the test,[1] but a higher frequency might improve the estimation of short run dynamics of the real exchange rate. But annual data are likely to be time averages - not end of period values, and Taylor (2001) shows that this will bias the coefficient estimate in an AR(1) model upwards. As a result, the half-life estimate will be biased upwards as well. Taylor (2001) argues that this feature might be one solution to the so-called PPP puzzle (Rogoff, 1996).[2] More recently, Imbs et al. (2005, forthcoming) have argued that aggregation across sectors may also cause an upward bias in the half-life estimates. But their work is challenged by Chen and Engel (2004).

On the other hand it has been known at least since the work of Orcutt (1948) that the OLS estimate of an AR(1) coefficient will be biased downwards. Median unbiased estimators that take this into account have been developed,[3] but just correcting for one type of bias when there are likely to be several can be misleading. In this case, for instance, it is not impossible that the different sources of bias will cancel. My approach will instead be to use standard methods and be aware of the potential biases when interpreting the results.

3.2 Real effective exchange rates

Another special feature with this study is that I have constructed real effective exchange rates (REER), that is, a weighted average of a number of bilateral real exchange rates.[4]

[1] I have discussed this in more detail in section 4.2.

[2] Zussman (2002) finds some evidence of this effect. Using annual, quarterly and monthly data from the same countries, he finds that the half-life estimates are higher the lower frequency of the data.

[3] See Cashin and McDermott (2003) for a study of PPP where such a method is applied.

[4] For a mathematical expression of a REER, see equation 3.2 and 3.3.

The huge majority of the PPP studies that have been done have instead chosen to focus exclusively on bilateral real exchange rates. The drawback with using REERs is that all the bilateral real exchange rates can be I(1), but a linear combination of them may for some reason be stationary, or cointegrated.[5] If a REER is found to be stationary, one cannot really claim to have found evidence of PPP.

On the other hand, for policy purposes, and in particular for an institution like the central bank of Norway, the most interesting variable is in fact the real effective exchange rate. If there is a tendency for the REER to revert to its mean over time, this is a relevant result for monetary policy purposes, even though it is debatable whether you can refer to it as PPP or not.

Another way to warrant the use of REERs is to imagine that the world only consisted of two countries - home and abroad. This is a frequent assumption in open economy models. Investigating the idea of PPP in such a setting necessitates the construction of a real effective exchange rate. The exercise would in any case not be that much different from what is done when studying the real exchange rate against the euro area these days. The exchange rate is the same in all countries, but when constructing price data, it will be a weighted average of price data from the euro members. This is in principle similar to what I do when I construct a weighted real exchange rate, with weights corresponding the countries' importance for Norway.

In any case I will also do a number of regressions on bilateral data.

The real exchange rate as compared with country i in a given year t is by definition given by

$$R_{it} = \frac{P_{it}^f}{P_t^h}, \tag{3.1}$$

where P_{it}^f is the price level in the foreign country and P_t^h the price level in the home country (Norway). This formula differs from the one introduced in section 2.1 because it assumes that the price levels are denoted in a common currency, which is the case in the Penn World Table. All price levels are expressed relative to the US price level in the current year and denoted in US dollars.

[5] I would like to thank Lucio Sarno for pointing this out to me.

The real effective exchange rate is constructed by Laspeyres' index formula. This formula is also used by the IMF and Norges Bank, among others, when constructing (real) effective exchange rates.

The real effective exchange rate in year t is given by

$$R_t = \prod_{i=1}^{n} {R_{it}}^{W_{it}}, \ where \ \sum_{i=1}^{n} W_{it} = 1. \tag{3.2}$$

In logs the same equation becomes

$$r_t = \sum_{i=1}^{n} W_{it} r_{it}. \tag{3.3}$$

The choice of weights may clearly influence the results. The optimal weights in the construction of a real effective exchange rate should correspond to the relative impact the respective countries have on the Norwegian economy. One argument for using import weights is that Norway is a small open economy, and that the majority of impulses from abroad come in the form of imported inflation. On the other hand, in order to maintain international competitiveness for the tradeable sector, this sector has traditionally played a leading role in the wage bargaining process in Norway. So even if Norway as a small open economy has no influence on prices abroad, it is not unlikely that developments in the most important export markets can have an impact on Norway. This warrants the use of trade weights instead of import weights.[6]

If export and import is relatively balanced versus most countries, this concern is superfluous. A quick inspection of the trade and import weights,[7] shows that the different weights are quite similar for most countries in most years, but that there are some exceptions. For instance, the weight assigned to the United Kingdom will in general be much higher if I use data on total trade instead of only import data.

But using trade data can also be criticized for being a too simplistic approach. Two

[6]The data that the trade and import weights are based on was obtained from Statistics Norway. The data are in current value in Norwegian kroner. The weight assigned to a specific country is the value of the import (trade) from (with) that country that year divided by the value of import (trade) from (with) all the countries in the respective group.

[7]The average weights over the post Bretton Woods period are reported in the appendix, table A.1.

countries that export the same type of commodities may not engage in an extensive, bilateral trade. But impulses from one of the countries are likely to be important for the other. This will not be captured by trade or import weights. The composition of a country's trade can also matter. If a country only exports goods that are priced on the world market, one would want to give it less weight than a country where the exporting firms enjoy some market power ("pricing to market").

These concerns have led to the development of far more sophisticated weighting measures. The IMF and the OECD have both constructed so-called competitiveness weights which are used in the construction of nominal and real effective exchange rates. These weights take the composition of the total trade into account and also include a measure of competition in third markets - not just bilateral trade flows.[8]

My main focus will nevertheless be on an importweighted real effective exchange rate. This is partially due to the fact that import weights were easiest available. In addition, even if the development in the most important export markets does matter, the impulses from the import markets are likely to be more consequential. But to check the robustness of my results, I have also constructed real effective exchange rates using total trade weights and the competitiveness weights of the OECD.

3.3 Econometric method and model

My null hypothesis will be that the real (effective) exchange rate is I(1). A rejection of the null will be interpreted as evidence in favour of PPP. Of course, that the real (effective) exchange rate is not I(1) is only a necessary condition for PPP. But since univariate regressions on single country data from the post Bretton Woods period generally have failed to reject the null hypothesis, I find this approach interesting enough.

Absolute PPP implies that the log of the real exchange rate has zero mean, since I have absolute price data here. As discussed in section 2.4, there are a number of reasons for why this condition might fail. Therefore, I will allow for a constant term. Rejection of the null

[8] A more detailed description and discussion of the IMF procedure is given in Zanello and Desruelle (1997). The OECD procedure is presented in Durand, Simon and Webb (1992).

hypothesis with a constant term significantly different from zero indicates that relative PPP holds. But I will generally not allow for a deterministic trend. Some authors argue that a real exchange rate model including a deterministic trend can be interpreted as a test of the Balassa-Samuelson modified version of PPP,[9] but it is not obvious that a deterministic trend is a good description of a process of relative productivity growth.

Another question is which testing procedure to apply to test for a unit root. A large number of different test statistics have been proposed, and, as discussed by Stock (1994), it is an inherent problem in unit root testing that it doesn't exist a univariate most powerful test of the autoregressive parameter being equal to 1 versus being less than 1 in a simple AR(1) model. Size and power vary substantially between different test statistics, and in addition, the properties of the test statistics can be very sensitive to the true nature of the model.

Faced with these difficulties I have chosen to employ the most frequently used test statistic, the augmented Dickey-Fuller t-statistic. This is an extension to account for serial correlation in the error terms of the traditional Dickey-Fuller test. But to give an indication of the robustness of the results, I will also report results from other test statistics. Monte Carlo simulations done by Stock (1994) indicate that the ADF t-statistic is good in terms of size, but that this comes at the cost of low power. The ADF t-test is a test of $\varrho = 1$ versus $\varrho < 1$ in (2.6).

An obvious problem is the choice of the number of lags k, and unfortunately, the results show that whether I can reject the null hypothesis or not is highly dependent on the number of lags included in the regression. Including too few may distort the size of the test, whereas including too many genereally reduces power (Ng and Perron, 1995)).

Numerous procedures have been proposed to select the appropriate number of lags. They range from deterministic rules that relate k to the number of observations[10] to data dependent methods that take sample information into account. An example of the latter is the Akaike information criterion (AIC), which is often reported automatically by econometric software programs.

[9] One example is Papell and Prodan (2004).

[10] An often used rule, $k = integer\ part\ of\ (12 \times (T/100)^{0,25})$, is due to Schwert (1989).

Simulation studies exploring the finite sample properties of the varying procedures are conducted by Hall (1994) and Ng and Perron (1995). They both find that the data dependent methods dominate the deterministic rules. Ng and Perron (1995) recommend a general to specific procedure with sequential F- or t-tests of the last lag(s) until the last lag(s) is found to be significantly different from zero. This is a valid procedure as long as the initial number of lags is chosen high enough. Data dependent selection rules such as the Bayesian information criterion (BIC) and AIC perform particularly poor when the true process contains a large, negative moving average component. The same authors (Ng and Perron, 2001) have later come up with a new testing procedure and an information criterion that performs better in unit root test, a modified AIC (MAIC). But since I am going to do a large number of regressions in this paper, I have decided to stick with sequential t-tests, since this approach is very simple to apply in PcGive 10, the program I used in the ADF regressions.

I have started with 4 lags for data series beginning in the 1950s ($k = 4$), 3 lags for series from the 1960s and 2 lags for series from the 1970s, and then eliminated lags until the last lag was found to be significant. This choice of the maximum number of lags might seem somewhat arbitrary, but because of the limited number of observations, I reasoned that I could not afford to lose more than two observations in the post Bretton Woods period. Anyway, it is only in a very few cases more than one lag is significant.

The other tests I have applied are the Phillips-Perron test (PP), due to Phillips and Perron (1988), the DF-GLS test developed by Elliott, Rothenberg and Stock (1996), and the KPSS test (Kwiatkowski et al., 1992). All these tests have been done using the econometrics program Eviews 5.0.

The PP test is, as the ADF test, designed to account for serial correlation in the residuals. But instead of adding lags in the initial regression equation, Phillips and Perron suggested to ignore the serial correlation and estimate the relation by OLS as an AR(1) model. This can be done since the OLS-estimate of the AR(1) coefficient is superconsistent[11] under the null hypothesis. They then proposed to modify the test statistic to account for the serial correlation.[12] Several procedures can be applied to do this. I have chosen to estimate the

[11] Converges in probability to the true value at a faster rate than required for consistency. See Hamilton (1994, p. 460).

[12] For a more pedagogical and detailed treatment of this test procedure, see for instance Hamilton (1994,

"correction factor" by using the Bartlett kernel, which is the standard procedure incorporated in the econometrics program Eviews 5.0. Unfortunately, one has to choose a bandwidth parameter exogenously for this estimator, and as with the choice of lags in the ADF-case, this choice can be decisive for the test conclusions, especially in small samples. I have chosen the bandwidth parameter by employing a data dependent rule due to Newey and West (1994), which is automatically computed in Eviews.

A third unit root test, which has been found to have both good size and power, is the DF-GLS test. This test is a modified version of the ADF test, and it involves two steps. First, the explained variable is regressed on a constant (and a trend, if that is included in the model), and then the usual ADF-procedure is run on the demeaned (or detrended) data, the residuals from the preceding regression. This procedure has been found to improve the asymptotic power of the test significantly in the presence of a unknown trend or constant, and simulation results show that the DF-GLS test performs well in small samples as well.[13] To determine the lag length in the DF-GLS tests I have used the MAIC, introduced by Ng and Perron (2001). This is automatically done in Eviews 5.0.

The fourth test procedure I have employed, the KPSS test, takes a different approach. The null hypothesis of this test is that the series is stationary - not non-stationary, as above. I will give a very simple example just to illustrate. A test of stationarity will be a test of the parameter $\theta = 1$ versus $\theta < 1$ in the following model

$$(1 - L)r_t = (1 - \theta L)\varepsilon_t,\ \varepsilon_t\ \sim NID(0, \sigma^2). \tag{3.4}$$

When $\theta = 1$, there is a common factor in $(1-L)r_t$ and $(1-\theta L)\varepsilon_t$, and r_t is stationary. Since existing unit root test all have flaws, it is often recommended to do stationarity tests as well, to get a cross-check of the results. I have therefore performed KPSS-tests on all regressions where there were significant rejection of the null hypothesis. Results are reported in the respective tables. As with the Phillips-Perron test, this test statistic requires a choice of the bandwidth parameter. I have also here employed the Newey-West bandwidth with the Bartlett kernel, which is automatically done in Eviews 5.0.

chapter 17.6)

[13]For Monte Carlo results, see Elliott, Rothenberg and Stock (1996)

Chapter 4

Empirical results

This chapter presents the empirical results of this paper. In the first section I provide a quick overview of the results from some other PPP studies done on Norwegian data. In the second section, I investigate the power properties of the DF t-statistic further, because some of the failures to detect evidence of PPP may be due to insufficient power. The third section contains estimation results for real effective exchange rates versus varying groups of countries. In the fourth section I look at several bilateral real exchange rates, including a series against the UK spanning almost two centuries. I also try to identify factors that can explain the variability in the persistence of deviations from PPP, measured against different countries. In the fifth and final section I try to check whether the real exchange rate is better described by a nonlinear model.

4.1 Previous studies

Earlier tests of PPP that have been done on Norwegian data have given mixed results. Bjørnland and Hungnes (2003), Chortareas and Driver (2001), Papell (1997) and Alexius (2001) all fail to reject the null hypothesis of the real exchange being I(1) at normal significance levels. But it deserves to be mentioned that most of these studies were done on short data sets, Bjørnland and Hungnes (2003), for instance, have data spanning 17 years only. Both Papell (1997) and Chortareas and Driver (2001) restrict their attention to the bilateral real exchange rate versus the United States.

On the other hand, Taylor (2002) is able to reject the null hypothesis employing a DF-GLS test on the Norwegian krone versus a world basket of currencies with more than a century of data. And Akram (2005, forthcoming) can reject the I(1) hypothesis both using an importweighted real effective exchange rate and bilaterally versus the United States, Germany and the United Kingdom. He uses a quarterly data set from 1972 to 2003.

Since lack of power is one possible explanation for the failure to reject the I(1) hypothesis, I will study the power properties of a frequently used unit root test, the Dickey-Fuller t-statistic, more closely in the following.

4.2 The power problem

Assume that the true process is a stationary AR(1) process, where the error term is white noise and the mean for simplicity is set to zero

$$r_t = \rho r_{t-1} + \varepsilon_t,\ |\rho| < 1,\ \varepsilon_t \sim NID(0, \sigma^2). \tag{4.1}$$

The DF t-statistic is given by

$$t_T = \frac{(\hat{\rho} - 1)}{\sigma_{\hat{\rho}}} = \frac{(\hat{\rho} - 1)}{(s_\varepsilon^2 / \sum\limits_{t=1}^{T} r_{t-1}^2)^{1/2}}, \tag{4.2}$$

where the denominator is the usual OLS standard error for the estimated coefficient. s_ε^2 is the OLS estimate of the residual variance

$$s_\varepsilon^2 = \sum_{t=1}^{T} (r_t - \hat{\rho} r_{t-1})^2 / (T - 1). \tag{4.3}$$

By using the asymptotic counterparts to the finite sample moments, I get the following formula for the t-statistic in a test where the null hypothesis is a random walk, as a function of the true parameter ρ and the sample size T (see the appendix, section A.3 for a formal derivation).

$$t_T^2 = \frac{T(\rho - 1)^2}{(1 - \rho^2)} \tag{4.4}$$

Table 4.1: Data span necessary to reject H_0

Half-life	Coefficient	Years necessary to reject H_0
1 year	0.5	25.1
2 years	0.70	48.7
3 years	0.79	72.6
12 months	0.94	24.1
24 months	0.97	48.2
36 months	0.98	72.3

The coefficient is found by solving (4.5) for ρ and inserting for different values of H. Equation 4.4 is then solved for T and the time span that is necessary for rejection is found by inserting for the autoregressive parameter ρ and the critical value t_T.

This approximation is clearly a simplification. But for practical purposes it only implies that it is likely that it is going to be even more difficult to reject the null hypothesis in finite samples.

The critical DF-value at a significance level of 5 percent is -2,89 (with 100 observations). In table 4.1 we have calculated the AR(1) coefficients for varying half-lives, and how many years of data that will be necessary to reject the hypothesis of nonstationarity, given that the real half-lives are as in the first column. The half-life is calculated by the usual formula,

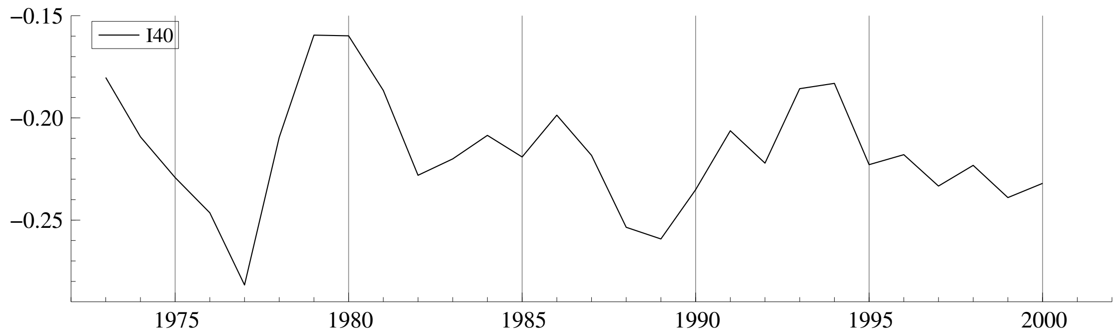
$$H = ln(0,5)/ln(\rho). \tag{4.5}$$

See the appendix, section A.2, for a more thorough discussion on how to calculate half-lives for various models. The number of years is found by solving (4.4) for T and inserting values for the critical value t_T and ρ.

With a true half-life of only one year, we would still need as much as 25 years of data before we can expect to be able to reject the null hypothesis. Since most of the PPP studies have reported a half-life of three years or more, it is unlikely that running regressions on data from one country spanning only the post Bretton Woods period will produce much evidence in favor of PPP.

A first attempt at solving the problem of low power might be to increase the frequency of the data, for instance, you might want to use monthly instead of annual data. This will lower

Figure 4.1: The Norwegian REER, I40



The graph shows the log of the Norwegian real effective exchange rate I40 over the in the post Bretton Woods period, 1973 to 2000.

the standard error of the estimator, but at the same time, the coefficient estimate approaches 1 and eliminates the gain. This feature has been known at least since the work of Shiller and Perron (1985). But there is some Monte Carlo-evidence that increasing the frequency might improve the modelling of short-run dynamics in finite samples.[1] More reasonable attempts at finding a solution to the power problem were discussed in section 2.3.2.

4.3 The Norwegian real effective exchange rate

The main focus in this paper is the Norwegian real effective exchange rate, constructed using import weights from 40 countries.[2] The countries are all included in I44, the importweighted effective exchange rate used by Norges Bank. I will refer to this series as I40. I look at data from 1973 to 2000. The data goes back to 1950 for quite a few countries, but I choose to concentrate on the post Bretton Woods period, as the world's major currencies have been floating against each other in this period. As I have discussed before (see section 2.3.2), it can be problematic to run regressions with data from periods with different exchange rate regimes. As will be seen later, it is also going to be much more difficult to reject that the real exchange rate is I(1) with data series starting in 1950.

The regression was specified without a trend, as the ADF model in (2.6), and a graphical

[1] See Choi (1992) for more on this issue.

[2] For a list of the countries included, see the appendix, section A.4.

inspection of I40 reveals that this is a reasonable specification (see figure 4.1). For completeness I ran the same regression allowing for a deterministic trend. The general results were the same. The estimated trend was negative (see table A.2 in the appendix), but not significantly different from zero in a t-test.[3]

Applying the general to specific procedure of sequential t-tests until the last lag is found to be significant, and allowing for a maximum of two lags, I decide to include one lag in the ADF regression. Alternative procedures of sequential F-tests of omitted lags or minimization of the AIC would have given the same conclusion in this case.

Turning to the results in table 4.2, we see that we have a surprisingly strong rejection of the unit root hypothesis. The null hypothesis is rejected even at a significance level of one percent. This conclusion holds with a considerably shorter sample. I ran recursive regressions, beginning in 1973 and ending in 1990 and onwards, and the unit root hypothesis was clearly rejected in all cases. But the rejections became stronger with more observations. The implied half-life of a unit shock over the period 1973 to 2000 is less than a year, 0,7 years calculated by the simple formula based on the coefficient in front of the level variable. See the appendix, section A.2 for more on this. If the lagged differenced term is omitted and the model estimated as an AR(1) model, the simple formula gives a half-life of about a year. Impulse response analysis, with all lags included, suggests a half-life of 1,4 years. See figure 4.2. Only having annual data is not optimal when we want to say something about the short run behaviour of the real exchange rate. The half-life estimates should therefore only be interpreted as an indication of the persistence of deviations from PPP. This is underlined by the fact that the three different methods gave quite different results. A 95 percent confidence interval around the ADF coefficient - ϱ in (2.6) - indicates that the coefficient is somewhere between 0,06 and 0,68 - corresponding to a half-life somewhere between one quarter and two years (0,25 years to 1,80 years).[4]

These results are in the same range as those reported by Akram (2005, forthcoming), on

[3] A t-test of the trend parameter being zero can only be done under the assumption that the equation does not contain a unit root, which I am testing here. But that assumption seems reasonable, given that the I(1) hypothesis is clearly rejected in this case.

[4] Again, this assumes that the true process is stationary. If the true coefficient is equal to one, the confidence intervals will be non-standard. A discussion of this is found in Stock (1994, pp. 2285-2288).

Table 4.2: ADF-equation for the Norwegian REER I40

$$\begin{array}{rcccccc} \Delta\hat{r}_t & = & -0,136 & - & 0,630\, r_{t-1} & + & 0,407\, \Delta r_{t-1} \\ & & (0,032) & & (0,151) & & (0,164) \end{array}$$

Estimation sample is 1973-2000, T=28

t-ADF = -4,16**, critical values(DF):	1%:	-3,685
R^2 = 0,423	5%:	-2,971
F(2,25) = 9,149[0,001]**	10%:	-2,625

Diagnostic tests

AR 1-2 test:	F(2,23)	=	1.7195 [0.2014]
ARCH 1-1 test:	F(1,23)	=	3.3054 [0.0821]
Normality test:	$Chi^2(2)$	=	0.3645 [0.8334]
hetero test:	F(4,20)	=	1.6272 [0.2065]
hetero-X test:	F(5,19)	=	1.2909 [0.3089]
RESET test:	F(1,24)	=	3.0719 [0.0924]

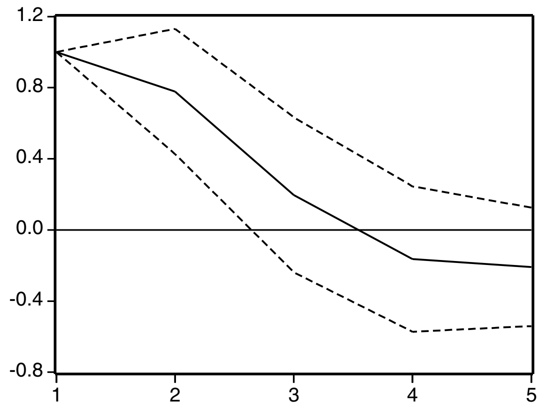
Estimation results for the natural log of the Norwegian real effective exchange rate I40 over the post Bretton Woods period. The numbers in parentheses are the standard errors of the estimated coefficients. The diagnostic tests reported are automatically computed by PcGive. The null hypotheses about the residuals are, from above, no autocorrelation, no conditional heteroskedasticity (ARCH), and normality, homogeneity and homogeneity including cross products. The regression specification test (RESET) is due to Ramsey (1969). See Hendry and Doornik (2001, chapter 18) for technical details about the tests. The numbers in the square brackets are the respective p-values. The sign ** implies significance at the 1% level here and in the rest of this paper. * and (*) imply significance at the 5% and the 1% level, respectively.

quarterly data on a real effective exchange rate from the post Bretton Woods period. His half-life estimate using the simple formula on the ADF-coefficient is less than four quarters, whereas impulse response analysis suggests a half-life of between 6 and 8 quarters.

Again assuming stationarity, we also note that a t-test of the constant term being zero is rejected. This means that PPP in the absolute form is rejected.[5]

[5]This result has interesting implications for the testing procedure. Given that the true model - not just the regression equation - contains a constant term, the asymptotic distribution of the DF t-statistic is in fact standard normal. This means that the DF critical values are too restrictive and in general will give

Figure 4.2: Impulse response graph for I40



The figure shows the impulse response function for a unit innovation for the estimated model for the REER I40. Time in years on the x-axis. The dotted lines represent a 95 percent confidence interval.

The diagnostic tests automatically reported in PcGive do not indicate that the regression equation is misspecified or any violation of the standard assumptions about the residuals (see table 4.2).

Since different unit root tests can give varying conclusions for the same data, I have also tested for a unit root in I40 using the Phillips-Perron Z_t test and the DF-GLS t-statistic (see table 4.6). Both methods reject the null hypothesis, but only at a significance level of five percent. In addition, stationarity is clearly accepted in a KPSS-test.

Another relevant question is whether the use of import weights is a good procedure, as discussed in section 3.2. In order to investigate this further, I have constructed real effective exchange rates against the same 14 countries, selected on the basis of available data, using import weights (IW), total trade weights (TW) and the competitiveness weights

too few rejections of the null hypothesis. But it is often recommended to use the DF critical values anyway because the true specification of the model is always uncertain. In addition, Banerjee et al. (1993) have suggested that the DF-distribution may be a better approximation than the standard normal distribution in finite samples.

Table 4.3: Correlation matrix for IW(14), TW(14) and CW(14)

	TW (14)	IW (14)	CW(14)
TW (14)	1.000	0.946	0.887
IW (14)	0.946	1.000	0.963
CW (14)	0.887	0.963	1.000

The correlation coefficients between the tradeweighted series, TW(14), the importweighted series, IW(14) and the series calculated by competitiveness weights, CW(14).

Figure 4.3: IW(14), TW(14) and CW(14) from 1973 to 2000

Real effective exchange rates for the same 14 countries based on trade weights (TW), competitiveness weights (CW) and import weights (IW) over the post Bretton Woods period.

(CW) calculated by the OECD.[6] [7] The correlation matrix between the three series are given in table 4.3, and the three series are graphed in figure 4.3. The formal analysis confirms the graphical impression of positive correlation. The correlation coefficient between the IW and TW series is 0,95, and 0,96 between the IW and CW series. The correlation between the CW and TW series is somewhat smaller.

It is a bit surprising that the ADF-test rejects the unit root hypothesis only for the TW

[6]When I selected 14 of the 25 countries in OECD's competitiveness weights for Norway, I rescaled the weights in order to make them sum to one. This is not a correct procedure, as the exclusion of one country in this index can change the relative relationship between the weights. But my hope is that the error from this step is not too significant.

[7]Source: OECD/Norges Bank. The weights are available at the webpage http://www.norges-bank.no/stat/valutakurser/weights.xls

Table 4.4: Test results for IW(14), TW(14) and CW(14)

Group	s.e.	t-ADF	DF-GLS	PP	KPSS	lags	hl(ADF)[8]
TW (14)	0,130	-3,532*	-1,990*	-2,570	0,546*	1	1,12
CW(14)	0,152	-2,705(*)	-2,064*	-2,197	0,546*	1	1,31
IW (14)	0,163	-2,726(*)	-2,143*	-2,424	0,550*	1	1,18

Table 4.5: Half-life for TW and CW

Group	lags	hl(ADF)	hl(AR1)	hl(IR)
CW24	1	0,82	1,40	1,7
TW16	1	1,08	1,80	2,2

series at a significance level of five percent, see table 4.4, but the two others are close to the critical values. When I employ the DF-GLS test, the unit root hypothesis is rejected for all the series. And if I include all the countries that I have data for in the respective series - not only the same 14 - the ADF-test rejects the unit root hypothesis at the one percent level for both the TW and the CW series. Results are reported in table 4.6. The implied half-life of deviations from PPP is about a year using the simple formula on the ADF-coefficient, but both the AR(1) approach and impulse response analysis suggest a half-life closer to 2 years. See table 4.5.

4.3.1 Subgroups

I40 is an aggregated variable of data on price levels and exchange rates from 40 countries. To get a better picture of the relationship with different countries, I have also constructed importweighted real effective exchange rates for subgroups of countries, grouped by geography and level of development. The countries included in the different groups are reported in section A.4 in the appendix.

We note that the null hypothesis of non-stationarity of the REER (see table 4.6) is rejected when compared to all subgroups of countries which are relatively close to Norway geographically. It is also rejected against subgroups of industrialized countries. It is in-

Table 4.6: Test for REERs from 1973, without a trend

Group	ADF-coef.	s.e.	t-ADF	lags	KPSS	DF-GLS	PP
S. Europe	0,856	0,104	-1,382	0	0,572*	-1,391	-1,268
Some to 1950	0,655	0,153	-2,251	0	0,542*	-2,129*	-2,436
Africa	0,856	0,090	-1,599	0	0,423(*)	-1,198	-1,719
Lat.-America	0,663	0,144	-2,339	0	0,177	-2,272*	-2,294
Less developed	0,826	0,105	-1,656	0	0,572*	-1,465	-1,508
Asia	0,649	0,150	-2,332	0	0,480*	-2,353*	-2,347
Scandinavia	0,553	0,142	-3,136*	1	0,132	-2,130*	-2,195
Nordic	0,478	0,153	-3,414*	1	0,111	-2,336*	-2,087
N. Europe	0,510	0,138	-3,546*	1	0,329	-2,264*	-2,079
Top ten	0,620	0,142	-2,671(*)	1	0,238	-1,922(*)	-1,988
Old industrial	0,525	0,151	-3,137*	1	0,496*	-2,189*	-2,480
Western	0,384	0,147	-4,192**	1	0,178	-2,439*	-2,926(*)
I40	0,370	0,151	-4,164**	1	0,155	-2,464*	-3,044*
CW24	0,428	0,149	-3,832**	1	0,487*	-2,435*	-2,423
TW16	0,527	0,127	-3,726**	1	0,466*	-2,058*	-2,397
Critical values		1%[**]	-3,685		0,739	-2,650	-3,689
		5%[*]	-2,971		0,463	-1,953	-2,972
		10%[(*)]	-2,625		0,347	-1,610	-2,625

teresting to observe that the DF-GLS test gives more rejections than the ADF-test, which again rejects more frequently than the Phillips-Perron test. As I have mentioned above, both asymptotic theory and Monte Carlo simulations have indicated that the DF-GLS test is the more powerful test.

When it comes to the ADF test, the asymptotic exercise with an AR(1) model, see section 4.2, indicated that with a true half-life of one year, we would need 25 years of data to reject the null hypothesis at a significance level of five percent. Here I have 28 years of data, and we observe that the null hypothesis only is rejected in the cases where the half-life estimate, calculated by the simple formula (see equation 4.5) on the ADF coefficient, is less than 1,2 years. This is to some extent a confirmation of the asymptotic calculations. I do in general find less support of PPP with subgroups based on non-European countries, and a graphical comparison with I40 reveals some of the reason, see figure 4.4. The variability is much larger, and some of the series appear to be trending downwards, indicating a long run appreciation

Table 4.7: Half-life for REERs from 1973, without a trend

Group	lags	hl(ADF)	hl(AR1)	hl(IR)	abs(IR-ADF)	abs(IR-AR1)
S. Europe	0	4,46	4,46	4,46	0	0
Some to 1950	0	1,64	1,64	1,64	0	0
Africa	0	4,46	4,46	4,46	0	0
Lat.-America	0	1,68	1,68	1,68	0	0
Less developed	0	3,62	3,62	3,62	0	0
Asia	0	1,60	1,60	1,60	0	0
Scandinavia	1	1,17	1,74	2,1	0,93	0,36
Nordic	1	0,94	1,49	1,8	0,86	0,31
N. Europe	1	1,03	1,77	2,1	1,07	0,33
Top ten	1	1,45	2,24	1,4	0,05	0,84
Old industrial	1	1,08	1,65	1,9	0,82	0,25
Western	1	0,72	1,11	1,6	0,88	0,49
I40	1	0,70	1,04	1,4	0,70	0,36
Average		1,89	2,19	2,29	0,41	0,23
Median		1,45	1,68	1,80	0,05	0,25
Maximum		4,46	4,46	4,46	1,07	0,84
Minimum		0,70	1,04	1,40	0,00	0,00

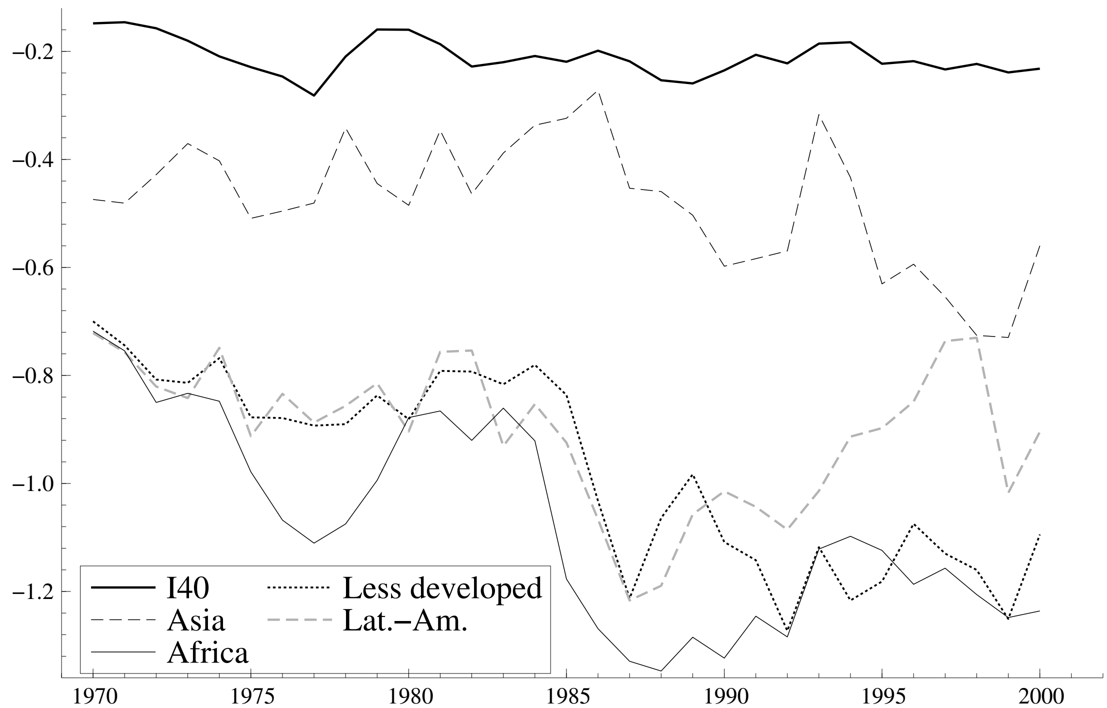
of the Norwegian real effective exchange rate. But allowing for a deterministic trend in the regression does not lead to more rejections.

For the series Southern Europe and Top Ten - Norway's top ten trading partners[9] - the inclusion of the trend changes the outcome of the ADF tests. Results are reported in table A.2 in the appendix. The unit root hypothesis is even rejected at the 1 percent level for Top ten, and the trend estimate is positive and significantly different from zero both for Top ten and Southern Europe.[10] We also note that the trend estimate is positive and significantly different from zero for several of the series where unit root behaviour was rejected without including a trend - implying a long run depreciation of the Norwegian real effective exchange rate.

[9]Based on average import from the respective countries in the period 1973 to 2000. I would in fact have picked the same countries if I based the selection on total trade data or OECD's competitiveness index, but the countries' relative importance would have been somewhat different.

[10]Again, the usual assumption of no unit roots in the series has to be done for the t-test of the trend parameter to be valid.

Figure 4.4: I40 and non-European REERs



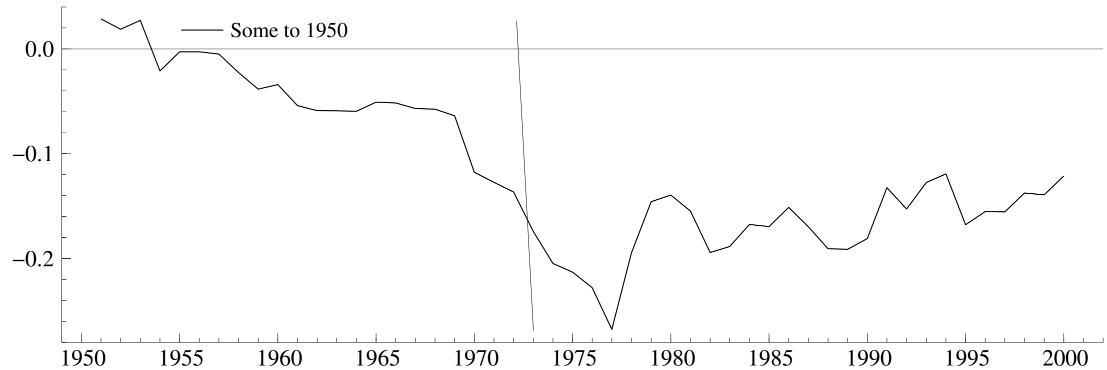
I40 compared with real effective exchange rates towards groups of non-European countries over the post Bretton Woods period. The countries included in the different groups are reported in the appendix, section A.4.

4.3.2 Longer REER time series

I have also constructed importweighted REER time series beginning before 1973 for countries where data was available. I have data back to 1960 for most countries, and back to 1950 for some. "Some to 1950" is the broadest long run REER series, constructed against 16 major trading partners. The complete list is given in the appendix, section A.4. A notable exception is Germany. The Penn World Table does not contain German data prior to 1970.

I do in general find it more difficult to reject the unit root hypothesis for data series spanning both the Bretton Woods period and the period after. The results are reported in table A.3 and table A.4 in the appendix. As already discussed, methodological concerns can be raised about running a regression on data from both a period with fixed exchange rates and a period where the world's major currencies have been floating. By doing this, I implicitly assume that the underlying dynamic process has been the same in the two periods.

Figure 4.5: The REER *Some to 1950*



A graphical inspection of the series "Some to 1950" reveals why (see figure 4.5). Measured against these countries, Norway apparently experienced a relatively strong real appreciation in the 1950s and 1960s. But if I restrict the attention to the post Bretton Woods period, stationarity does not seem to be an incorrect description. As we remember (see table 4.6), the unit root hypothesis was rejected for this series using the DF-GLS test. If we follow the reasoning of the basic PPP hypothesis, a gradual liberalization of world trade should lead to stronger convergence to PPP. And increasing trade liberalization is exactly what has characterized the development in international trade since 1950. Thus, the idea that PPP is more valid now than before sounds reasonable.

But this storyline misses an important point - that nominal exchange rates generally were fixed under Bretton Woods, whereas they have been floating afterwards. Real exchange rate variability have generally been higher after Bretton Woods, mainly due to large swings in the nominal exchange rate. Obstfeld and Taylor (2004, p. 116) report average half-lifes of deviations from PPP for a number of countries in different periods, and the average half-life is somewhat lower in the Bretton Woods period than in the period after. My results do not fit into this picture.

Another concern here that I think is relevant is the reliability of the data. The first benchmark study that is the basis for the Penn World Table was done in 1970, and the earlier data are based on extrapolations using data from national accounts. Their reliability is therefore likely to decrease with the distance from 1970. This is another reason for why I have chosen to concentrate on the post Bretton Woods period in this study.

4.4 Bilateral real exchange rates

The majority of the published PPP studies have been done on bilateral data. In order to be able to compare my findings with those studies, I have also performed unit root tests and estimated implied half-life individually for all countries in my data set. This exercise may give some hints on whether the surprisingly low half-life estimate for I40 is due to swift mean reversion also bilaterally. The alternative explanation is that Norway in general will experience a real appreciation against some other countries, and a real depreciation versus others. These opposing effects might roughly cancel, and as a result, a real effective exchange rate will be more stationary than bilateral real exchange rates.

I have done regressions both starting in 1950 and 1973, with and without a deterministic trend. As usual, the main focus will be on the post Bretton Woods period, without allowing for a deterministic trend. All the results are reported in tables A.5 to A.9 in the appendix, and all the bilateral real exchange rates over the period 1973 to 2000 are graphed at the end of the appendix.

My results reproduce the picture given by many other PPP studies. It is difficult to reject the I(1) hypothesis on bilateral real exchange rates with data only for the post Bretton Woods period. At the usual significance level of five percent, I can only reject the random walk hypothesis against four out of 40 countries using the standard ADF test. Surprisingly enough, one of them is Argentina. Studying a graph of the Norwegian-Argentinian real exchange rate for the relevant period, see figure A.1 in the appendix, we see that it in fact appears to be stationary around a constant, but the deviations from the "equilibrium" level are exceptionally large. The stationarity is in any case probably more of a coincidence than due to strong goods market arbitrage between Norway and Argentina. The other rejections are more expected from a PPP perspective - Germany, Netherlands and Iceland. We also note that we can reject the unit root hypothesis for Belgium, Brazil (?) and the US at the ten percent level.

As before I obtain more rejections of nonstationarity with the DF-GLS test - this test procedure gives some evidence of PPP also for France, Sweden and the Philippines (?), whereas the PP test as usual produces fewer rejections than the ADF. The KPSS test of stationarity results in rejections for 25 of the 40 real exchange rates I consider. The KPSS test does not reject stationarity in any of the cases where non-stationarity was rejected .

Allowing for a deterministic trend gives some new rejections of unit roots on the post Bretton Woods period. Some of the previous rejections now become insignificant, as the inclusion of the trend seems superfluous and results in a loss of power. And as with the real effective exchange rates, it is more difficult to reject the I(1) hypothesis with data series from 1950 and onwards.

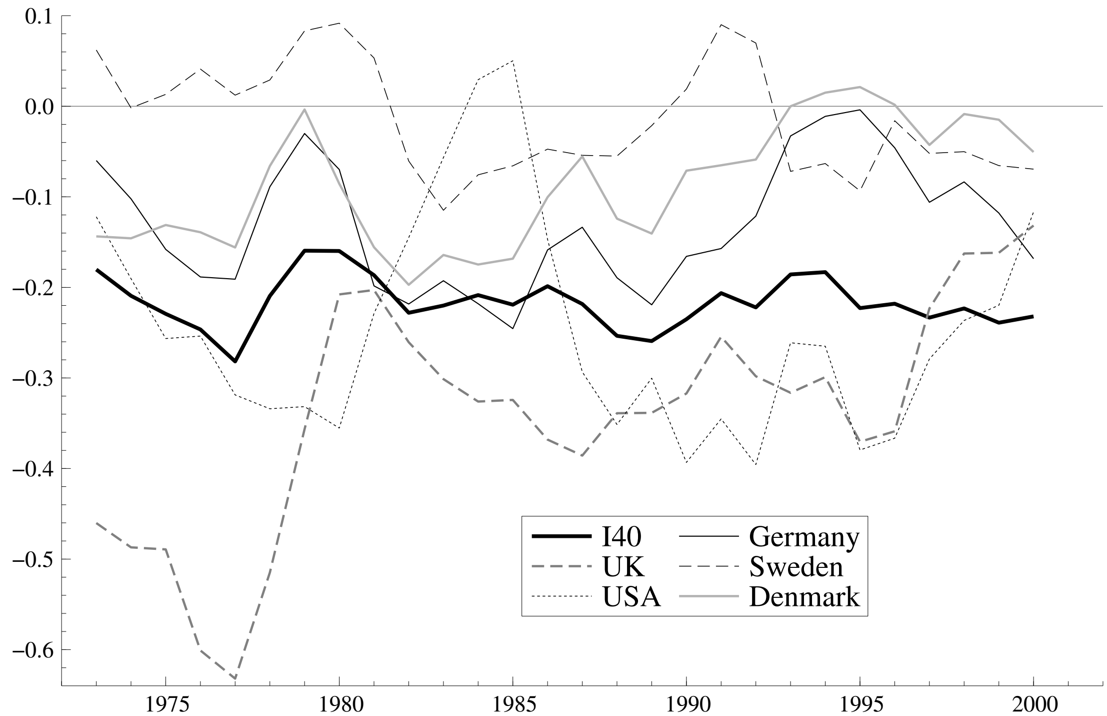
Before I conclude this section, it is worthwhile spending some more time on the estimated half-lives for the post Bretton Woods regression without a constant term. The average half-life is 3,9 years, based on the simple formula (equation 4.5) on the ADF-coefficient. The AR(1) half-life average is 4,4 and the impulse response average 4,3. This is within the 3-5 year consensus view reported in the literature. The median half-life (ADF) is a little bit less, 2,6 years. And for no countries is the estimated half-life less than 0,68 years (Iceland). This is practically equal to the estimated half-life (ADF) for the broadest REER, I40 (0,69). This indicates that I40 displays relative fast mean reversion because opposing shocks roughly cancel - not because some individual mean reversion rates are even faster than for the REER, as suggested by Akram (2005, forthcoming).[11] A graphical comparison (see figure 4.6) of I40 and some important bilateral real exchange rates strengthens this hypothesis.

4.4.1 UK-Norwegian real exchange rate, 1819 to 2003

Through the project "Historical Monetary Statistics for Norway 1819 - 2003", which was published in October 2004 (Eitrheim, Klovland and Qvigstad, 2004), Norges Bank has constructed a consumer price index beginning in 1519 and gathered exchange rate data from 1819 and onwards. One outcome of this project was the construction of a bilateral real

[11] This comparison can be a bit misleading since a REER is a *weighted* average of bilateral real exchange rates - not just a regular average where each country has the same weight. Therefore I constructed a weighted average of the bilateral half-lives as well, using the average import weights over the post Bretton Woods period as weights. The weighted averages were a bit shorter - 2,6 years for the ADF half-life, 3,1 for the impulse response half-life and 3,2 for the AR(1) half-life. But my conclusion - that the REER has a considerably faster speed of convergence to PPP than bilateral real exchanges rates - is unchanged.

Figure 4.6: I40 and major bilateral real exchange rates

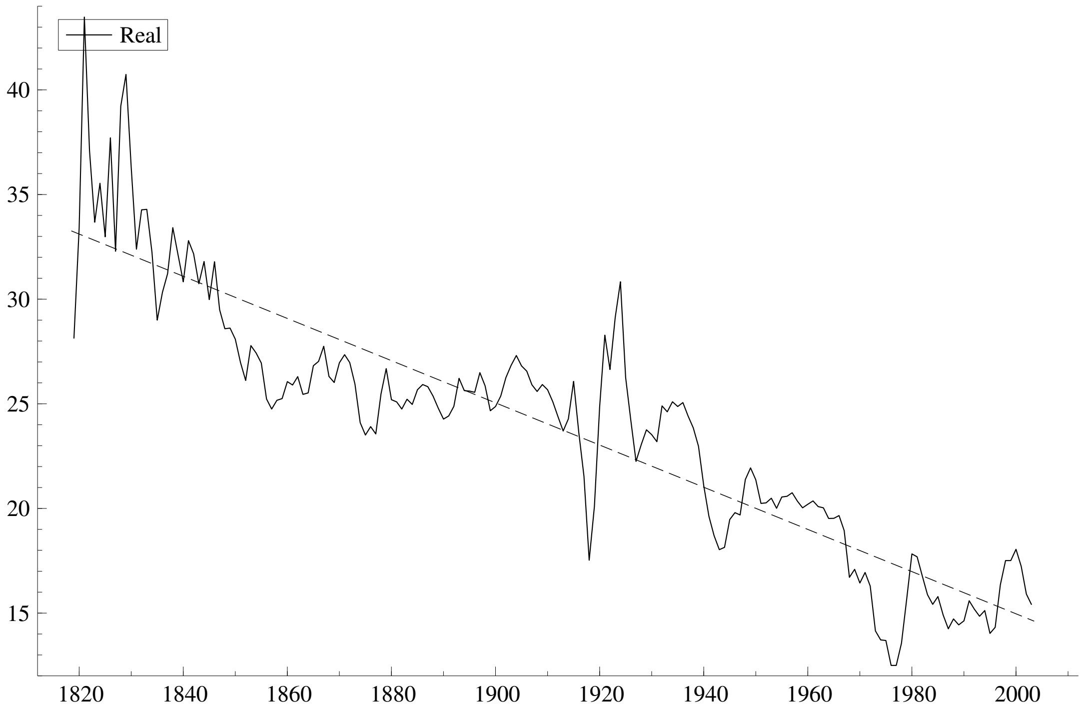


exchange rate towards the UK for the period from 1819 to 2003[12].

Edison and Klovland (1987) have previously analysed the UK-Norwegian real exchange rate over the period 1874 to 1971 and found evidence of PPP if they allowed for different short run dynamics from 1914 to 1928. But a PPP analysis on such a long time series as I have here, has never been done on Norwegian data before. As I have discussed before in this paper, to do a simple PPP analysis with data from periods with different exchange rate regimes, is a questionable procedure. But this will not intimidate us from making an attempt, when the data are already available. The long run UK-Norwegian real exchange rate series is graphed in figure 4.7. As we can see, the series is clearly not stationary around a constant level. This is confirmed by the standard unit root tests, reported in table 4.8. The various tests have been performed in the same way as previously in this paper (see section 3.3). But if I allow for a deterministic trend, the conclusions change. Non-stationarity is now clearly

[12]I would like to thank Øyvind Eitrheim for allowing me to use this time series. Details on how the series was produced and sources are given in Eitrheim, Klovland and Qvigstad (2004, chapter 7)

Figure 4.7: The UK-Norwegian real exchange rate, 1819-2003



The figure shows the UK-Norwegian real exchange rate over the period 1819 to 2003. The dotted line is the OLS-estimated trend.

Table 4.8: Test results for the UK-Norwegian real exchange rate

Specification	ADF-coef.	lags	t-ADF	DF-GLS	PP	KPSS	Trend
Without trend	0,959	0	-1,855	-1,434	-1,289	1,557**	–
With trend	0,742	0	-5,307**	-3,219*	-5,465**	0,138(*)	-0,027*

rejected by all the unit root tests, and the ADF-test (here equivalent to a DF-test, since no lagged differenced terms were found to be significant in a t-test) and the PP test even reject at the one percent level. The DF-GLS test rejects at the five percent level only, and the reason for that is probably that the modified AIC, which I have used to determine the number of lags for this statistic, points to three differenced terms. But the null hypothesis is rejected irrespective of how many lags I include. I have allowed for a maximum of four lags, which is a reasonable assumption since I have annual data - that the real exchange rate more than four years ago should influence the real exchange rate today sounds unlikely.

The estimated trend is negative and clearly significantly different from zero in a t-test. The implied half-life of deviations from the negative trend is found to be 2,3 years. The three methods to calculate half-life as I have described before give the same result here, as no differenced lagged terms are included in the model. Diagnostic tests reported by PcGive does not indicate any residual autocorrelation or misspecification (RESET). But there are evidence of conditional heteroskedasticity (ARCH) and usual heteroskedasticity in the residuals. Normality is also rejected. This undermines the reliability of the tests, but the graph of the UK-Norwegian real exchange rate does not indicate that trend stationarity is an incorrect conclusion. To save space, the diagnostic tests are not reported .

The results imply a long run appreciation of the Norwegian real exchange rate. If we believe in the Balassa-Samuelson effect, this can be explained by a relative productivity growth advantage in the tradable sector for Norway compared with the UK. Productivity growth cannot be directly observed, but the growth in real GDP per capita has sometimes been used as a measure of this. Therefore, I have gathered data series starting in 1830 on real GDP for Norway and the UK.[13] According to these data, Norway has on average experienced an annual growth in real GDP per capita on 2,1 percent from 1831 to 2000. Over the same period the annual average growth for the UK was 1,4 percent. This fits well with the Balassa-Samuelson effect and a long run appreciation of the Norwegian real exchange rate. Decomposing the real GDP data for different periods I find that the annual growth on average was about the same for the UK and Norway in the 19th century (1,3 and 1,2 percent respectively), but that real GDP for Norway grew considerably faster in the 20th century. From 1901 to 1950 the average is 0,9 percent for the UK and 2,3 percent for Norway, and in the second part of the century, Norway's growth was 3,1 percent on average, whereas the UK grew with 2,2 percent.

If this story is correct, this should translate into a relatively stable real exchange rate in the 19th century and a real appreciation in the 20th century. If the period from 1819 till about 1850 is excluded, when the real exchange rate volatility according to the data was exceptionally large, the real exchange rate appears to be pretty stable until World War I and the abandonment of the gold standard. After World War I there is no doubt that the

[13] My sources are, for Norway, Eitrheim, Klovland and Qvigstad (2004), and for the UK, Officer (2003).

long run tendency is an appreciation.

A next step here would be to try model explicitly the real exchange rate and real GDP growth - or perhaps an even better measure of productivity growth in the tradable sector. The Balassa-Samuelson theory would be strengthened if these variables turn out to be stationary around a constant mean. But that step belongs in a different paper.

4.4.2 What explains half-life?

The estimated half-lives are a measure of the speed of convergence to PPP compared with different countries. Of course, this is only a very rough measure, and it needs to be interpreted with caution. In addition, I have not been able to reject the unit root hypothesis vis-á-vis the large majority of the countries, and if the series are in fact non-stationary, it is misleading to estimate half-lives.

But given that I have already calculated the half-lives, it can be interesting to investigate whether I can identify factors that explain the variations. The basic reasoning for PPP is goods market arbitrage. We would therefore expect that the half-life towards Norway's main trading partners would be shorter, everything else equal. To get a measure of this, I have taken an average of the relative import weights in the period from 1970 to 2000.

Another possible explanation is physical distance, as transportation costs are likely to increase with distance. In this sense larger obstacles to trade can pave the way for larger deviations from PPP. For a measure of this, I have collected data on "greater circle distance" between the capital cities for all the countries.[14] A third possible factor is the level of development. If we take the Balassa-Samuelson effect into account, then in the case of Norway, it would be more natural to expect that PPP would hold against countries that have been continuously among the world's most developed. To get a measure of this, I have taken an average of gross domestic product per capita in current prices over the post Bretton Woods period. The data are from the Penn World Table. I have also taken data on the degree of openness, in current prices, from the PWT, to check whether this might

[14] My source is http://www.wcrl.ars.usda.gov/cec/java/capitals.htm. This source is also used by Zussman (2002) as a measure of distance between countries.

Table 4.9: Half-life and explanations

Variable	a)	b)	c)	d)	e)
Distance	4,8E-04 (4,4E-04) [0,28]	8,9E-04 (3,6E-04 [0,019]	_	_	_
Distance^2	-4,2E-08 (3,1E-08) [0,18]	-6,5E-08 (2,7E-08) [0,021]	_	_	_
CGDP	-1,2E-04 (1,1E-04) [0,28]	_	-2,0E-04 (7,9E-05) [0,017]	_	_
Import	-10,549 (13,665) [0,45]	_	_	-20,505 (11,083) [0,072]	_
OPENC	-6,8E-04 (8,1E-03) [0,93]	_	_	_	-0,003 (0,008) [0,724]
R^2	0,213	0,141	0,142	0,083	0,003

Regression results from OLS-regressions with half-life (ADF) as dependent variable. The symbol "_" indicates that the respective variable was not included in the regression. The numbers in paratheses ( ) are the standard errors of the estimates. The numbers in the square brackets [ ] are the respective p-values. The variable CGDP is per capita GDP in current prices, OPENC is a measure of openness in current prices. The respective regressions are numbered from a) to e).

help explain the half-life variation.[15] We would expect half-life to be shorter for more open economies.

I then run an OLS-regression with half-life as the dependent variable, and all the above mentioned variables as explanatory variables on the right hand side. I have also included the squared distance as an explanatory variable, since it is natural to expect that transportation costs will be an increasing, but concave function of distance. We would therefore expect a negative coefficient for the squared distance. This coefficient estimate does indeed have the expected sign, as can be seen in table 4.9, and the situation is the same for all the other variables. But none of them are significantly different from zero in a standard t-test.

[15]The openness variable in PWT is simply total trade (the sum of exports and imports) divided by gross domestic product.

Table 4.10: Correlation of explanatory variables

	Import	CGDP	Distance	OPENC
Import	1,000	0,519	-0,438	-0,090
CGDP	0,519	1,000	-0,406	0,247
Distance	-0,438	-0,406	1,000	0,035
OPENC	-0,090	0,247	0,035	1,000

One possible explanation for this is multicollinearity among the explanatory variables. We would expect both per capita GDP and average import weights to be negatively correlated with distance, as the world's most developed countries are close to Norway, and Norway's main trading partners are the neighbouring countries. This is confirmed by studying the correlation matrix of the explanatory variables (table 4.10). We also note that per capita GDP and average import weights are positively correlated. The openness variable appears to be more independent. This is the only variable that is clearly not significantly different from zero in a t-test when all the other explanatory variables are omitted from the regression. It therefore seems appropriate to conclude that openness - at least in the way as it is measured here - does not help explain the half-life variability.

Given that the other explanatory variables clearly are correlated, it might be that they capture much of the same effects. We note that the estimated coefficients are larger in absolute value in the individual regressions. For my purposes, the good thing about distance between capitals is that this is clearly an exogenous variable. So if it is not relevant in itself, it might function as an instrumental variable for either per capita GDP or average import. We note that in the regression with only distance, the coefficient on the first order term is positive, and the coefficient on the squared term is negative, as expected. They are both significantly different from zero in t-tests, but a joint F-test of both coefficients being zero simultaneously cannot be rejected, although it is very close, with a p-value of 0,06. Diagnostic tests of the residuals did not indicate that the standard assumptions of homogeneity and no correlation were violated, but normality was clearly rejected in all the regressions. This undermines the reliability of the tests I performed.

I repeated this exercise with half-life calculated by the impulse response function and and half-life based on an AR(1) model as dependent variables. The results were roughly the same.

A more thorough study of the determinants of half-life variability would use higher frequency data and preferably look at more than one country. There are also other variables that might be included, like cultural and linguistical similarities, preferential trade agreements and better measures of trade costs than just distance. But until such a study is done, a preliminary conclusion is that distance/trade costs, average import share and per capita GDP are possible explanations of half-life variability.

4.5 Nonlinearities?

A recent strand of the PPP literature has tried to solve the PPP puzzle by applying nonlinear models to the real exchange rate. As discussed in section 2.3.2, the existence of transaction costs give reasons to believe that the real exchange rate will become increasingly mean reverting the further away it is from its mean.

In the literature, two different nonlinear models have been been used on the real exchange rate. One is the threshold autoregressive model. An example of this is Zussman (2002). The TAR-model implies that the real exchange rate can be non-stationary within a band, but that it becomes mean reverting once it crosses one of the thresholds. Zussman finds the half-life of deviations from the edge of the band to be 1,18 years, whereas a standard Dickey-Fuller regression (without a band of inaction) on the same data set gives a half-life estimate of 3,7 years.

But since the real exchange rate is an aggregated variable, the rationale for distinct thresholds as in the TAR model can be questioned. Intuitively one would expect that the dynamic process is more gradual. Taylor, Peel and Sarno (2001) try to capture this feature by using a so-called exponential smooth transition autoregressive model (ESTAR), and their half-life estimates are well below the consensus estimates obtained using linear models. For large shocks they do for instance find a half-life of less than one year for the dollar-sterling and dollar-mark real exchange rate.

The mathematical formula for the ESTAR model as employed by Taylor, Peel and Sarno is the following

$$r_t - \mu = \sum_{i=1}^{k} \beta_i (r_{t-i} - \mu) + \sum_{i=1}^{k} \beta_i^* (r_{t-i} - \mu) \left(1 - exp[-\theta^2 (r_{t-d} - \mu)^2]\right) + \varepsilon_t, \tag{4.6}$$

where the integer $d > 0$ is a delay parameter. The appealing feature of the ESTAR model is that it allows for symmetric adjustment above and below the mean, which we would expect of a real exchange rate. It can also capture different speeds of reversion depending on the distance from the mean. To decide whether the real exchange rate follows a nonlinear process, and if so, which model to apply, Taylor, Peel and Sarno followed the procedure described in Teräsvirta (1994). He outlines a test of nonlinearity where the null hypothesis is that the process is linear against the alternative hypothesis of a smooth transition autoregressive model (STAR).

The procedure involves three steps. The first step is to specify a linear regression model, including the number of lags. Then the delay parameter d is determined by testing linearity for different values of d. If linearity is rejected, the third step is to choose between an ESTAR and a LSTAR (logistic) model. This is done by performing a sequence of F-tests.

As emphasized by Teräsvirta, Tjøstheim and Granger (1994), the fact that linearity is rejected does not necessarily imply that a nonlinear model will be a better choice than a linear model. The class of nonlinear models is infinite, and linearity tests are usually designed to have power against a wide range of alternatives. In small samples, the test methods are inclined to overfit, and may thus incorrectly reject linearity. Some types of heteroscedasticity may also be confused with nonlinearity in the tests. As a result of this, Teräsvirta, Tjøstheim and Granger (1994) recommend to compare the models out-of-sample and choose the one that performs best there.

This last step is omitted in Taylor, Peel and Sarno (2001), but they have checked the robustness of their results in a number of other ways.

I have followed in the footsteps of Taylor, Peel and Sarno and performed linearity tests on the Norwegian real exchange rate versus Sweden, Denmark, the United Kingdom and the United States. The countries were chosen because they are among Norway's most important trading partners but not part of the euro area. This provides us with data from 1973 and until today without obvious structural breaks, which is an advantage if I am going to do out-of-sample evaluation of the estimated model.

Previously in this paper I have used annual data from the Penn World Table. But when I am going to distinguish between linear and nonlinear models, and perhaps estimate by nonlinear techniques, a large number of observations are necessary. I have therefore gathered monthly data on consumer prices and exchange rates (end-of-month values) from the IMF's IFS database. The estimation sample is from January 1973 to December 2000. The remaining observations are retained for post sample evaluation. I have performed the tests with real exchange rate data both in the original form and transformed to natural logs.

In the specification of the linear model (4.6 without the nonlinear terms), the number of lags, k, is found to be only one. This means that I have an AR(1) model. I have tried with a large number of information criteria (Akaike, Schwarz, Hannan-Quinn and their modified versions), t-tests of the last lag and inspection of the partial autocorrelation function, and they all arrive at only one lag. This is the case both using the original values and the natural logs. The hypothesis of no autocorrelation in the residuals is also clearly accepted.

The next step is to choose the delay parameter d. The recommendation in Teräsvirta (1994) is to perform linearity tests with different d's and choose the d that gives the smallest p-value. I have tried with d's ranging from one to six. The results are reported in table 4.11 with the smallest value in **bold** letters. For the original values, we note that linearity is only rejected at the 5 percent level for the US real exchange rate, although both Denmark and Sweden are very close to being significant. The d value that gives the smallest p value for the US real exchange rate is 5. For the natural log values, the US real exchange rate is again the only real exchange rate where linearity is rejected. This time the minimizing d value is 3. Denmark is close to rejection, whereas linearity is clearly accepted for Sweden and the UK. It deserves to be mentioned here that normality of the residuals is clearly rejected for all my series. This undermines the reliability of the tests.

The third step, given that linearity is rejected, is to choose between a LSTAR model and an ESTAR model. For the US real exchange rate with original values and a d-parameter of 5, the sequential F-tests suggest that a LSTAR model gives the best fit to the data (see table 4.12). I just note in passing that with $d = 1$ as the delay parameter, the tests would

Table 4.11: Determination of the delay parameter d

Country	lags(k)	p(d=1)	p(d=2)	p(d=3)	p(d=4)	p(d=5)	p(d=6)
USD	1	*0,044*	0,131	0,029	0,026	**0,004**	0,102
DKK	1	0,368	0,763	**0,052**	0,073	0,173	0,236
SEK	1	0,235	**0,057**	0,085	0,398	0,729	0,756
GBP	1	**0,157**	0,901	0,699	0,774	0,748	0,737
usd	1	0,025	0,022	**0,004**	0,005	0,020	0,274
dkk	1	0,490	0,449	0,080	**0,062**	0,118	0,362
sek	1	0,325	**0,153**	0,176	0,395	0,544	0,568
gbp	1	0,666	0,470	0,580	0,580	0,618	**0,284**
usd-gbp	1	0,259	0,199	0,167	0,072	**0,016**	0,085
usd-jpy	1	0,057	**0,009**	0,010	0,023	0,163	0,143

The table reports results from the application of the method Teräsvirta (1994) prescribed for determining the delay parameter d. His recommendation was to choose the d that gave the lowest p-value in the linearity test. The respective p-values here are given in **bold** letters. The tests were performed on bilateral real exchange rates, with Norway as a base country, against the United States (USD), Denmark (DKK), Sweden (SEK) and United Kingdom (GBP). Uppercase letters are original values, lowercase letters represent natural logs. The two last entries are the bilateral real exchange rate (in logs) between the US and UK, and the US and Japan, the real exchange rates Taylor, Peel and Sarno (2001) studied.

have indicated an ESTAR model. Linearity was nearly rejected for Sweden and Denmark, and the same tests suggest a LSTAR model also in these cases.

For log values, the tests again indicate that an LSTAR model is the most appropriate for the US real exchange rate. For Denmark, the tests now point to an ESTAR model with a delay parameter of 4, but in this case linearity was not rejected at a significance level of five percent.

For the real exchange rates Taylor, Peel and Sarno (2001) studied, they found that the delay parameter was equal to one in all cases. In addition, the tests unambiguously pointed in the direction of an ESTAR model. Since my results with Norway as a base country clearly differed from those of Taylor, Peel and Sarno (2001), I redid their study with data from the same source (IFS) and the same time period (January 1973 to December 1996). I considered the real exchange rate between the US and the UK and the US and Japan. With data transformed to logs, a d parameter of 5 gave the lowest p-value for the US-UK real exchange

Table 4.12: Model selection, ESTAR or LSTAR?

	H01 (p)	H02 (p)	H03 (p)	Model
USD	0,001	0,306	0,268	LSTAR
USD	*0,658*	*0,006*	*0,565*	*ESTAR*
DKK	0,048	0,684	0,057	LSTAR
SEK	0,007	0,724	0,638	LSTAR
GBP	0,748	0,371	0,547	ESTAR
usd	0,002	0,082	0,319	LSTAR
dkk	0,869	0,007	0,853	ESTAR
sek	0,366	0,035	0,964	ESTAR
usd-gbp	0,006	0,230	0,253	LSTAR
usd-jpy	0,005	0,600	0,061	LSTAR

The table contains the p-values from the tests described by Teräsvirta (1994) to distinguish between an ESTAR model and a LSTAR model. If the p-values of H01 and H03 are less than H02, it points to a LSTAR model. If the p-value of H02 are relatively smaller, it indicates an ESTAR model. The row in *italics* are the p-values for the real exchange rate versus the US with the delay parameter d set to 1. Uppercase letters are original values, lowercase letters represent natural logs

rate. The sequential tests gave no clear answer, but pointed towards a LSTAR model. For the US-Japanese real exchange rate, I found d to be two, and the tests again indicated a LSTAR model. My results are reported in the tables 4.11 and 4.12. The only reason why my results differed that I can think of is that there must have been some revisions in the IFS database from the time Taylor, Peel and Sarno (2001) gathered their data and up till today.

Of course, the LSTAR and ESTAR models can be close substitutes in many cases. But for my purposes, the only interesting model is the ESTAR model. The LSTAR model fits best to a process with switching between two regimes, and does not have the ESTAR-property of symmetric adjustment above and below an equilibrium, which can be explained by the story of transaction costs and gradual convergence to PPP.

Anyway, the results only gave significant nonlinearities in the Norwegian real exchange versus the US. Just for the sake of having tried it, I made the ad hoc assumption that the d-parameter could not be be more than two (which is reasonable given that I only found one lag to be significant in the AR model). With this assumption, linearity is rejected with a d parameter equal to one, and the sequential tests suggest an ESTAR model. I tried to estimate the model (4.6) by nonlinear least squares with these values for d and k. Eviews

5.0 returned reasonable parameter estimates, consistent with the hypothesis that the real exchange rate will exhibit faster mean reversion the further away it is from its mean, but unfortunately for my purposes, they were clearly not significantly different from zero. The results are not reported here.

So even if Taylor, Peel and Sarno (2001) made some progress on the road to a "solution to the PPP puzzles", the results indicate that their solution is in no way general. Another issue is that with the relatively fast speed of mean reversion I have found for the Norwegian real effective exchange rate in this paper, Rogoff's PPP puzzle may not be that relevant for Norway.

Chapter 5

Concluding remarks

The main finding in this paper was that I40, an importweighted real effective exchange rate constructed for Norway against a group of 40 countries, was clearly not integrated of order one over the period 1973 to 2000. Thus, the theory of relative purchasing power parity seems to describe the behaviour of I40 well over this period.[1] At the same time, I found it much more difficult to reject the I(1) hypothesis for bilateral real exchange rates against the same 40 countries. One possible explanation is that a REER will be more stationary than bilateral real exchange rates because opposing shocks in different countries will roughly cancel. The graph of I40 and some major bilateral real exchange rates in figure 4.6 is not inconsistent with this explanation. More support is provided by the fact that the estimated half-life of deviations from PPP for I40 is considerably shorter than the average half-life bilaterally against the same 40 countries. This is the case irrespective of which method I employ to calculate the half-life.

The huge majority of the PPP studies have been done on bilateral data, to some extent because it is debatable whether the detection of stationarity for REERs can be interpreted as evidence of PPP. This was discussed in section 3.2. A natural question to ask next is whether REERs in general tend to be more stationary than bilateral real exchange rates. Cashin and McDermott (2003) study the REERs constructed by the IMF for 20 industrial countries. Their data spans the post Bretton Woods period. They calculate half-lifes based on various estimators, but when they estimate the relation straightforwardly as an AR(1)

[1] Unsurprisingly, we found no evidence of absolute PPP.

model, the average half-life is found to be about five years, which is in the upper end of the consensus reported in the literature for bilateral real exchange rates. Accordingly, that study does not indicate that REERs in general display faster mean reversion than bilateral real exchange rates. Interestingly enough, the half-life for Norway in that study is found to be only 22 months, well below the average. Only Iceland displays even faster mean reversion among the 20 industrial countries. Other countries with short half-lives are New Zealand and Switzerland.

The results reported by Cashin and McDermott (2003) indicate that the relatively fast speed of mean reversion I have found for the REER I40 is a feature that distinguishes Norway from many other countries. This also suggests that the PPP theory has been more relevant for Norway than for many other countries. This is in some sense counterintuitive, as many economists would expect that the oil discoveries and the extensive public spending that was financed by the oil revenue would result in a long run real appreciation of the currency - and divergence from PPP.

Two cross-country studies that has tried to identify factors that explain the variability of persistence of deviations from PPP are Cashin and McDermott (2004) and Cheung and Lai (2000). Unfortunately, these two studies do not agree on much. Cashin and McDermott's conclusions are based on the REERs constructed by the IMF, whereas Cheung and Lai use US-based bilateral real exchange rates. This is one possible explanation for why their results differ.

But they share a few conclusions. Both studies agree that countries with higher inflation rates have shorter half-lives, and that productivity growth and trade openness do not seem to explain much. Cheung and Lai identify a high level of government spending as a factor contributing to longer half-lives, but Cashin and McDermott do not find this factor to be significant. None of this is very helpful in accounting for the relatively fast mean reversion of the Norwegian REER I40, as Norway typically has had low inflation and is characterized by openness to trade and a relatively high level of government spending - which according to these studies only should give a longer half-life.

Akram (2005, forthcoming) bases his argument on Cheung and Lai's finding of quick mean reversion for developing countries when he tries to explain the short half-life of Norway in his study. As for many developing countries, Norway's export mainly consists of nonmanufactured goods. But Cashin and McDermott challenge Cheung and Lai's somewhat unexpected result, arguing that it is due to a biased selection of countries. Cashin and McDermott instead detect a significantly shorter half-life for industrial countries than for developing countries.

Another factor that can be relevant is the exchange rate regime. The world's major currencies have been floating against each other in the post Bretton Woods period, but Norway's policy in the period has been to stabilize the nominal currency against its main trading partners. Whether nominal exchange rate stability contributes to real exchange rate stability in general, is unclear. Cheung and Lai's answer to this question would be positive, whereas Cashin and McDermott find that half-lives on average are longer for countries with pegged currencies. They also identify a negative relationship between a measure of nominal exchange rate variability and half-life. A story that can explain this feature, is that a more variable nominal exchange rate will adjust more easily to the level consistent with PPP if inflation rates differ.

But it might be that a cross-country study can give misleading conclusions here. If a country has a huge debt denoted in foreign currency, the authorities will often maintain a fixed parity even if the currency is clearly overvalued, because a devaluation will increase the burden of the debt. This is a phenomenon that has predominantly been observed in developing countries. It illustrates that a fixed exchange rate regime can function quite differently depending on a country's debt. Some historical studies of industrial countries point to faster mean reversion with a fixed exchange rate. Obstfeld and Taylor (2004) found considerably shorter half-life in the Bretton Woods period than in the period after with a group mainly consisting of industrial countries. Chowdhury and Sdogati (1993) studied PPP among major European countries under the European Monetary System (EMS). Even with only 13 years of data, they detected evidence of PPP bilaterally among major EMS countries, but not relative to the US. The authors also failed to find support of PPP before the EMS was introduced.

I will therefore argue that the fact that Norway has actively tried to stabilize the nominal exchange rate against its main trading partners after the collapse of the Bretton Woods system, is likely to have contributed to the relative stability of the real exchange rate since then. It has also been an explicit Norwegian policy to try to maintain the competitiveness of the tradable sector. This has for instance been reflected in the way the wage bargaining process has been organized, largely in line with the Scandinavian model of inflation. The main intention of the frequent devaluations in the 1970s and 80s was also to preserve the viability of the tradable sector. A third example it the so-called "solidarity alternative" ("solidaritetsalternativet") in the 1990s, a coordinated attempt at restricting the wage growth. A related factor that can help us understand the fast mean reversion is that Norway has shown less real wage rigidity than other industrial countries. This is usually ascribed to Norway's centralized wage bargaining system. See Akram (2005, forthcoming) and Akram (2002, chapter 5.3) for a short, but still more detailed description of special features of the Norwegian economy.

Summing up this discussion we may conclude that Rogoff's famous PPP puzzle does not seem to be very relevant for Norway, at least not when we are looking at a real effective exchange rate. But why this is the case, is still another puzzle from my point of view, although I have mentioned some possible explanations above.

Other findings in this paper was that the level of development, distance and the level of trade are possible explanations of variability in bilateral half-life of deviations from PPP. But since these factors were clearly correlated, it might be the case that they have captured much of the same effect. As we remember, none of the estimated parameters were significantly different from zero when all the variables were included in the regression simultaneously.

I also found the UK-Norwegian real exchange rate over nearly two centuries to be trend stationary, with a negative trend. This implies a long run real appreciation of the Norwegian exchange rate, which is consistent with the Balassa-Samuelson effect, as Norway has seen a considerably faster growth in real GDP than UK in the relevant period.

Other studies have found that some real exchange rates are characterized by nonlinear mean reversion, and that this can be one solution to Rogoff's PPP puzzle. But the tests I performed on some bilateral Norwegian real exchange rates did not indicate any meaningful nonlinearity.

Bibliography

Abuaf, N. and P. Jorion (1990). Purchasing power parity in the long run. *Journal of Finance 45*(1), 157–174.

Akram, Q. F. (2002). State dependent effects in labour and foreign exchange markets. Doctoral dissertations in economics 1, Norges Bank.

Akram, Q. F. (2005, forthcoming). PPP in the medium run despite oil shocks: The case of Norway. *Journal of Macroeconomics*.

Alexius, A. (2001). Sources of real exchange rate fluctuations in the Nordic countries. *Scandinavian Journal of Economics 103*(2), 317–331.

Balassa, B. (1964). The purchasing power parity doctrine: A reappraisal. *Journal of Political Economy 72*(6), 584–596.

Banerjee, A., J. Dolado, J. W. Galbraith, and D. F. Hendry (1993). *Co-integration, Error-correction, and the Econometric Analysis of Non-stationary Data*. Oxford: Oxford University Press.

Banerjee, A., J. D. Dolado, D. F. Hendry, and G. W. Smith (1986). Exploring the equilibrium relationships in econometrics through static models: Some Monte Carlo evidence. *Oxford Bulletin of Economics & Statistics 48*(3), 253–278.

Bhagwati, J. N. (1984). Why are services cheaper in the poor countries? *Economic Journal 94*, 279–386.

Bjørnland, H. C. and H. Hungnes (2003). Fundamental determinants of the long run real exchange rate: The case of Norway. Discussion Paper 326, Statistics Norway.

Cashin, P. and C. J. McDermott (2003). An unbiased appraisal of purchasing power parity. *IMF Staff Papers 50*(3), 321–51.

Cashin, P. and C. J. McDermott (2004). Parity reversion in real exchange rates: Fast, slow, or not at all. Working Paper 04/128, IMF.

Chen, S.-S. and C. Engel (2004). Does aggregation bias explain the PPP puzzle? Manuscript, University of Wisconsin, Madison.

Cheung, Y.-W. and K. S. Lai (2000). On cross-country differences in the persistence of real exchange rates. *Journal of International Economics 50*, 375–397.

Choi, I. (1992). Effects of data aggregation on the power of tests for a unit root: A simulation study. *Economics Letters 40*, 397–401.

Chortareas, G. E. and R. L. Driver (2001). PPP and the real exchange rate-real interest differential puzzle revisited: Evidence from non-stationary panel data. Working Paper 138, Bank of England.

Chowdhury, A. R. and F. Sdogati (1993). Purchasing power parity in the major EMS countries: The role of price and exchange rate adjustment. *Journal of Macroeconomics 15*(1), 25–45.

Cumby, R. E. (1996). Forecasting exchange rate changes and relative prices with the hamburger standard: Is what you want what you get with McParity? Working Paper 5675, NBER.

De Gregorio, J., A. Giovannini, and H. C. Wolf (1994). International evidence on tradables and nontradables inflation. *European Economic Review 38*(6), 1225–1244.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*, 1161–1176.

Dornbusch, R. (1985). Purchasing power parity. Working Paper 1591, NBER.

Dornbusch, R. and P. Krugman (1976). Flexible exchange rates in the short run. *Brookings Paper on Economic Activity*, 647–668.

Durand, M., J. Simon, and C. Webb (1992). OECD's indicators of international trade and competitiveness. Working Paper 120, OECD.

Edison, H. J. and J. T. Klovland (1987). A quantitative reassessment of the purchasing power parity hypothesis: Evidence from Norway and the United Kingdom. *Journal of Applied Econometrics 2*(4), 309–333.

Eitrheim, Ø., J. T. Klovland, and J. F. Qvigstad (Eds.) (2004). Historical monetary statistics for Norway 1819-2003. Occasional Paper 35, Norges Bank.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996). Efficient tests for an autoregressive unit root. *Econometrica 64*, 813–836.

Engel, C. and J. H. Rogers (1996). How wide is the border? *The American Economic Review 86*(5), 1112–1125.

Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica 55*(2), 251–276.

Frankel, J. A. (1986). International capital mobility and crowding out in the US economy: Imperfect integration of financial markets or goods markets? In R. Hafer (Ed.), *How open is the US economy?* Washington, DC: American Enterprise Institute.

Frenkel, J. A. (1978). Quantifying international capital mobility in the 1980s. In D. Bernheim and J. Shoven (Eds.), *National Saving and economic performance*. Chicago: University of Chicago Press.

Frenkel, J. A. and H. G. Johnson (1976). *The Monetary approach to the balance of payments.* London: Allen & Unwin.

Froot, K. A., M. Kim, and K. Rogoff (1995). The law of one price over 700 years. Working Paper 5132, NBER.

Froot, K. A. and K. Rogoff (1991a). The EMS, the EMU, and the transition to a common currency. Working Paper 3684, NBER.

Froot, K. A. and K. Rogoff (1991b). Government consumption and the real exchange rate: The empirical evidence. Mimeo, Harvard Business School.

Froot, K. A. and K. Rogoff (1995). Perspectives on PPP and long-run real exchange rates. In G. Grossmann and K. Rogoff (Eds.), *Handbook of International Economics*. Amsterdam: North Holland Press.

Geary, R. C. (1958). A note on exchange rates and purchasing powers between countries. *Journal of the Royal Statistical Society 121*, 89–99.

Hall, A. (1994). Testing for a unit root in time series with pretest data-based model selection. *Journal of Business & Economic Statistics 12*(4), 461–470.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Hendry, D. F. and J. A. Doornik (2001). *Empirical Econometric Modelling Using PcGive*. London, UK: Timberlake Consultants Ltd.

Heston, A., D. A. Nuxoll, and R. Summers (1994). The differential-productivity hypothesis and purchasing-power parties: Some new evidence. *Review of International Economics 2*(3), 227–243.

Heston, A., R. Summers, and B. Aten (2002). Penn World Table version 6.1. Center for International Comparisons at the University of Pennsylvania (CICUP). http://pwt.econ.upenn.edu/.

Imbs, J., H. Mumtaz, M. O. Ravn, and H. Rey (2005, forthcoming). PPP strikes back: Aggregation and the real exchange rate. *Quarterly Journal of Economics*.

Johnson, D. R. (1990). Co-integration, error correction and purchasing power parity between Canada and the United States. *The Canadian Journal of Economics 23*(4), 839–855.

Kasa, K. (1992). Adjustment costs and pricing-to-market: Theory and evidence. *Journal of International Economics 32*(1-2), 1–30.

Kilian, L. and M. P. Taylor (2003). Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics 60*(1), 85–107.

Knetter, M. M. (1997). Why are retail prices in Japan so high? Evidence from German export prices. *International Journal of Industrial Organization 15*(5), 549–572.

Kravis, I. B. and R. E. Lipsey (1983). Toward an explanation of national price levels. Princeton Studies in International Finance 52, Princeton University.

Krugman, P. R. (1987). Pricing to market when the exchange rate changes. In S. W. Arndt and J. D. Richardson (Eds.), *Real-financial linkages among open economies*, pp. 49–70. Cambridge, Mass. and London: MIT Press.

Krugman, P. R. (1990). Equilibrium exchange rates. In W. H. Branson, J. A. Frenkel, and M. Goldstein (Eds.), *International policy coordination and exchange rate fluctuations*, pp. 159–187. University of Chicago Press.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics 54*(1-3), 159–178.

Lothian, J. R. and M. P. Taylor (1996). Real exchange rate behavior: The recent float from the perspective of the past two centuries. *The Journal of Political Economy 104*(3), 488–509.

Mark, N. C. (2001). *International Macroeconomics and Finance. Theory and Econometric Methods.* Malden, USA; Oxford, UK: Blackwell Publishers.

Meese, R. A. and K. Rogoff (1988). Was it real? The exchange rate-interest differential relation over the modern floating-rate period. *The Journal of Finance 43*(4), 933–948.

Mussa, M. (1986). Nominal exchange rate regimes and the behaviour of real exchange rates: Evidence and implications. *Carnegie Rochester Conference Series on Public Policy 25*, 117–214.

Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance matrix estimation. *The review of Economic Studies 61*, 631–653.

Ng, S. and P. Perron (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association 90*(429), 268–281.

Ng, S. and P. Perron (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica 69*(6), 1519–1554.

Obstfeld, M. and K. Rogoff (1995). The intertemporal approach to the current account. In G. M. Grossman and K. Rogoff (Eds.), *Handbook of international economics*, Volume 3, pp. 1731–1799. Amsterdam: North Holland Press.

Obstfeld, M. and K. Rogoff (1996). *Foundations of International Macroeconomics*. Cambridge, Massachusetts: The MIT Press.

Obstfeld, M. and A. M. Taylor (1997). Nonlinear aspects of goods-market arbitrage and adjustment: Heckscher's commodity points revisited. *Journal of the Japanese and International Economies 11*(4), 441–479.

Obstfeld, M. and A. M. Taylor (2004). *Global Capital Markets. Integration, Crises and Growth*. Cambridge, UK: Cambridge University Press.

Officer, L. H. (2001). Exchange rate between the United States dollar and the British pound, 1791-2000. Economic History Services, EH.Net. URL: http://www.eh.net/hmit/exchangerates/pound.php.

Officer, L. H. (2002). What was the gold price then? Economic History Services, EH.Net. URL: http://www.eh.net/hmit/goldprice/.

Officer, L. H. (2003). The annual real and nominal GDP for the United Kingdom, 1086 - 2000. Economic History Services, EH.Net. URL: http://www.eh.net/hmit/ukgdp/.

Oh, K.-Y. (1996). Purchasing power parity and unit root tests using panel data. *Journal of International Money and Finance 15*(3), 405–418.

Orcutt, G. H. (1948). A study of the autregressive nature of the time series used for Tinbergen's model of the economic system of the United States, 1919-1932. *Journal of the Royal Statistical Society, Series B 10*(1), 1–45.

Papell, D. H. (1997). Searching for stationarity: Purchasing power parity under the current float. *Journal of International Economics 43*, 313–332.

Papell, D. H. and R. Prodan (2004). Long run purchasing power parity: Cassel or Balassa-Samuelson. Working Paper, University of Houston.

Parsley, D. C. and H. A. Popper (2001). Official exchange rate arrangements and real exchange rate behavior. *Journal of Money, Credit and Banking 33*(4), 976–993.

Phillips, P. C. B. and P. Perron (1988). Testing for a unit root in time series regression. *Biometrika 75*, 335–346.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least squares analysis. *Journal of the Royal Statistical Society B 31*, 350–371.

Rogoff, K. (1996). The purchasing power parity puzzle. *Journal of Economic Literature 34*, 647–668.

Roll, R. (1979). Violations of purchasing power parity and their implications for efficient international commodity markets. In M. Sarnat and G. Szego (Eds.), *International Finance and Trade*. Cambridge, Massachussets: Ballinger.

Rossi, B. (2003). Confidence intervals for half-life deviations from purchasing power parity. Unpublished paper, Duke University, Department of Economics.

Samuelson, P. A. (1964). Theoretical notes on trade problems. *The Review of Economics and Statistics 46*(2), 145–154.

Sarno, L. and M. P. Taylor (2002). Purchasing power parity and the real exchange rate. *IMF Staff Papers 49*(1), 65–105.

Schwert, G. W. (1989). Tests for unit roots: A Monte Carlo investigation. *Journal of Business & Economic Statistics 7*(2), 147–159.

Shiller, R. J. and P. Perron (1985). Testing the random walk hypothesis: Power versus frequency of observation. *Economics Letters 18*(4), 381–386.

Stock, J. H. (1994). Unit roots, structural breaks and trends. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume IV. Elsevier Science.

Stone, R., D. G. Champernowne, and J. E. Meade (1942). The precision of national income estimates. *The Review of Economic Studies 9*(2), 111–125.

Summers, R. and A. Heston (1991). The Penn World Table (Mark 5): An expanded set of international comparisons, 1950-1988. *The Quarterly Journal of Economics 106*(2), 327–368.

Taylor, A. M. (2001). Potential pitfalls for the purchasing power parity puzzle? Sampling and specification biases in mean-reversion tests of the law of one price. *Econometrica 69*, 473–498.

Taylor, A. M. and M. P. Taylor (2004). The purchasing power parity debate. Working Paper 10607, NBER.

Taylor, M. P. (1988). An empirical examination of long-run purchasing power parity using cointegration techniques. *Applied Economics 20*(10), 1369–1381.

Taylor, M. P. (2004). Is official exchange rate intervention effective? *Economica 71*(281).

Taylor, M. P., D. A. Peel, and L. Sarno (2001). Nonlinear mean reversion in real exchange rates: Towards a solution to the purchasing power parity puzzles. *International Economic Review 42*, 1015–1042.

Teräsvirta, T. (1994). Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association 89*(425), 208–218.

Teräsvirta, T., D. Tjøstheim, and C. W. Granger (1994). Aspects of modelling nonlinear time series. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume IV. Elsevier Science.

Zanello, A. and D. Desruelle (1997). A primer on the IMF's information notice system. Working Paper 97/71, IMF.

Zussman, A. (2002). The limits of arbitrage: Trading frictions and deviations from purchasing power parity. Discussion Paper 02-12, Stanford Institute for Economic Policy Research.

Appendix

A.1 The Penn World Table (PWT)

The Penn World Table (version 6.1), produced by the Center for International Comparisons at the University of Pennsylvania, is a data set of national accounts time series covering 168 countries for some or all of the years from 1950 to 2000. The PWT provides the best measure of absolute price levels across countries that is available today.

The PWT is constructed with a basis in benchmark studies done by the International Comparison of Prices Program (ICP) of the United Nations. The program was initiated in 1968, and the first benchmark study covering 16 countries was conducted two years later. Since then worldwide studies covering an increasing number of countries have been done in 1975, 1980, 1985 and 1996. For the period 1985 to 1996 a number of regional studies were done, some of them on an annual basis (the EU members). From 1980 and onwards Eurostat, the OECD, the World Bank and UN's regional offices have been involved in the program.

Gross domestic product is divided into 150 categories, and national prices for 400 to 700 items are typically collected. A very relevant concern here is of course if the items are indeed equal. In order to address this concern, specification manuals containing detailed descriptions of the items have been developed. In some cases it is necessary to use professionals to set the right price. This is for instance done in construction, where architects are involved in the pricing process.

A particularly difficult task is the pricing of services, since many services are not priced in a market. This is of course equally difficult for the national statistical authorities. The ICP has adopted the common procedure of pricing a service at the value of the inputs used.

After the item prices have been collected, the items are grouped by category and the item prices are expressed as ratios of the corresponding item prices in the US. The price ratios in each category are then averaged, obtaining 150 price parities for each country. This last step is done since not all item prices are available for every country. The individual price parities are then aggregated based on a procedure first suggested by Geary (1958). The price level aggregate is expressed in US dollars.

The remainder of the work - the construction of the Penn World Table - is then done by the Center for International Comparisons at the University of Pennsylvania. Since the benchmark surveys are not done every year for every country, price data from national accounts are used to connect the studies. Quite often this procedure results in a mismatch between different benchmark studies. A simple errors in measurement model is used to reconcile ICP benchmark comparisons and the national accounts growth rates. A priori information about the reliability of the benchmark and national accounts data are exploited in this process. This way of solving the connection problem follows the frequently used procedure in national accounting developed by Stone, Champernowne and Meade (1942).

As indicated above, the Penn World Table contains data on a large number of countries that have never been subject of a benchmark study. The price levels of these countries were estimated based on capital price surveys done by the UN's International Civil Service Commission. These surveys are done in order to equalize real incomes of high-ranking executives assigned to different countries. The quality of the data from these countries are of course questionable, but as I have mentioned before, there are no better measure of the absolute price level available.

In the appendix to the PWT, the publishers have included a quality grading of the data from the various countries, with the grades ranging from A to D. In my sample, I40, 16 countries received the A grade, 9 were given a B and 15 were rated C. Norway was by the way given an A. The grades for the respective countries are reported in table A.6. More information about the PWT and how the data have been produced can be found in Summers and Heston (1991).

A.2 Calculation of half-lives

The question of how to calculate half-lives of deviations from PPP and their corresponding confidence interval has actually been the topic of several papers in the literature. See Rossi (2003) and references therein. Given that the true process is stationary, the estimated half-life is a measure of the persistence of shocks. In many cases I have not been able to reject the unit root hypothesis, and if the true model contains a unit root, calculating half-lives is misleading. But it can still be interpreted as a measure of persistence. We would of course expect the half-life estimates to be higher for non-stationary series.

If I assume that the real exchange rate follows an AR(1) process with zero mean as in equation 4.1, then, at horizon h, the percentage deviation from equilibrium is ρ^h. The half-life of deviation from PPP is defined as the smallest value of h such that

$$E(r_{t+h}|r_{t-s}, s \leq 0) \leq \frac{1}{2}r_t, \tag{A.1}$$

where E is the expectation operator. That is:

$$\rho^h = \frac{1}{2} \Rightarrow h = \frac{ln(1/2)}{ln(\rho)} \tag{A.2}$$

So if the estimated model is an AR(1) model, the calculation of half-lives are straightforward. But problems arise if the estimated model is an AR(p) model, with p larger than one. In the literature it is not uncommon to ignore the serial correlation in the residuals and estimate the model as an AR(1) model.[2] The resulting half-life estimate is clearly an approximation only. Another approximation[3] is to use the estimated coefficient (ϱ) in front of the level variable in the ADF regression (see equation 2.6), and insert this value for ρ in (A.2). This procedure can give a reasonable half-life estimate if one of the roots in the model is much closer to one than the others. The largest root will dominate the dynamic process, and the coefficient in front of the level variable is an approximation to this root. But as can be seen from this discussion, it is only under strict assumptions that using the level coefficient from the ADF regression is a good approximation. A better procedure is to use the impulse response function (IR) based on the dynamic multipliers (see Hamilton

[2]Abuaf and Jorion (1990) is an example of this.

[3]See Mark (2001, chapter 2.4.)

(1994, chapter 1.1-2)), and to calculate half-lives by solving $IR(H) = \frac{1}{2}$ for H. But since this procedure can be computationally intensive, researchers have sometimes reported the approximations above instead.

I have computed half-lives based on the ADF approximation for all the regressions I have done, and the results are reported in the respective tables. But for the regressions done on the post Bretton Woods period I have also calculated half-lives based on the coefficient from an AR(1) regression on the same data. In addition, I have calculated half-lives from the impulse response function by assuming that the process is linear for non-integer values. This is of course a simplification.

From the results on bilateral data on the post Bretton Woods period, see table A.6 in the appendix, we se that the average half-life is about the same for the three methods. The ADF approach gives an average half-life of 3,9 years, the AR(1) average is 4,4 years and the impulse response average is 4,2 years. It must be mentioned here that no lags are found to be significant in many of the regressions, and in these cases, the three methods give identical results. That the three averages are relatively close to each other is therefore no surprise. We notice that the AR(1) average is closer to the IR average than the ADF-average, and the average absolute deviation from the IR results is also slightly smaller for the AR(1) half-lives (0,34 compared to 0,41). If I assume that the IR method is the optimal method for calculating half-lives, the results therefore indicate that the AR(1) approximation is better than the ADF approximation.

A.3 Derivation of (4.4)

σ_i^2 denotes the theoretical (and asymptotic) variance, $i = r,\ \varepsilon,\ \hat{\rho}$, whereas s_i^2 is the empirical variance. The symbol $\rightarrow$ does here imply convergence in probability.

$$s_\varepsilon^2 = \sum_{t=1}^{T} (r_t - \hat{\rho} r_{t-1})^2/(T-1) \rightarrow \sigma_\varepsilon^2 \tag{A.3}$$

by the law of large numbers, since $\hat{\rho}$ is a consistent estimate of ρ.

Equivalently,

$$s_r^2 \rightarrow \sigma_r^2 = \frac{\sigma_\varepsilon^2}{(1-\rho^2}. \tag{A.4}$$

I then proceed to the squared t-statistic

$$t_T^2 = \frac{(\hat{\rho}-1)^2}{s_{\hat{\rho}}^2} = \frac{(\hat{\rho}-1)^2}{(s_\varepsilon^2 / \sum_{t=1}^{T} r_{t-1}^2)} = \frac{(\hat{\rho}-1)^2\, T s_r^2}{s_\varepsilon^2}. \tag{A.5}$$

By inserting the true value instead of the estimated $\hat{\rho}$, and inserting the asymptotic variances from (A.3) and (A.4), I arrive at the expression from (4.4),

$$t_T^2 \to \frac{(\rho-1)^2\, T \frac{\sigma_\varepsilon^2}{(1-\rho^2)}}{\sigma_\varepsilon^2} = \frac{T(\rho-1)^2}{(1-\rho^2)}. \tag{A.6}$$

A.4 Groups of countries

This sections reports the countries that are included in the various groups used in this paper.

I44 Argentina, Austria, Australia, Bangladesh, Belgium, Brazil, Canada, Switzerland, Chile, China, Colombia, Germany, Denmark, Spain, Finland, France, United Kingdom, Greece, Hong Kong, Hungary, Indonesia, Ireland, India, Iceland, Italy, Japan, South Korea, Sri Lanka, Morocco, Malaysia, Netherlands, Philippines, Pakistan, Portugal, Sweden, Singapore, Thailand, Turkey, USA, South Africa; *Czech Republic, Poland, Russia and Taiwan.*

I40 The same as I44, but the four countries mentioned at the end (in *italic letters*), are excluded because the Penn World Table doesn't have data on them from the entire period 1973 to 2000.

Western Austria, Belgium, Canada, Switzerland, Germany, Denmark, Spain, Finland, France, United Kingdom, Greece, Ireland, Iceland, Italy, Netherlands, Portugal, Sweden, Turkey, USA.

Asia Bangladesh, China, Hong Kong, Indonesia, India, Japan, South Korea, Sri Lanka, Malaysia, Philippines, Pakistan, Singapore, Thailand (Singapore is excluded from the series starting in 1960).

Top ten Germany, Denmark, France, Finland, United Kingdom, Italy, Japan, Netherlands, Sweden, USA (based on average import weight in the period 1973 to 2000.)[4]

Less developed Argentina, Bangladesh, Brazil, Chile, China, Colombia, Hong Kong, Indonesia, India, South Korea, Sri Lanka, Morocco, Malaysia, Philippines, Pakistan, Singapore, Thailand, Turkey, South Africa (Singapore excluded from the series starting in 1960).

Old industrial Austria, Australia, Belgium, Canada, Switzerland, Germany, Denmark, Spain, Finland, France, United Kingdom, Greece, Ireland, Iceland, Italy, Japan, Netherlands, Portugal, Sweden, USA (Germany excluded form the series starting in 1960).

S. Europe (Southern Europe): Spain, Greece, Italy, Portugal and Turkey.

N. Europe (Northern Europe): Austria, Belgium, Switzerland, Germany, Denmark, Finland, France, United Kingdom, Ireland, Iceland, Netherlands, Sweden (Germany excluded from the series starting in 1960).

Nordic Denmark, Finland, Iceland, Sweden.

Scandinavia Denmark, Sweden.

Some to 1950 Austria, Belgium, Canada, Switzerland, Denmark, Spain, Finland, France, United Kingdom, Iceland, Italy, Japan, Netherlands, Portugal, Sweden, USA.

Lat.-Am. (Latin America): Argentina, Brazil, Chile, Colombia

Africa Morocco, South Africa.

IW(14), TW(14), CW(14) USA, Japan, France, Italy, UK, Canada, Austria, Belgium, Denmark, Finland, Netherlands, Spain, Sweden and Switzerland.

[4] I would get the same group of countries if I used total trade weights or competition weights for the same period, but their relative importance would be somewhat different.

A.5 Empirical results

The following tables reports the standard error (s.e) of the level coefficient in an ADF-regression (ϱ in 2.6) and the t-value in an ADF-test of a unit root. The KPSS column is the KPSS test statistic. Remember that the null hypothesis here is that the series are stationary, not non-stationary as the other unit root tests. I also report the estimated trend coefficient if that is included in the regression, the number of lagged differenced terms included (the value of k in equation 2.6) and the number of observations for each series. DF-GLS is the DF-GLS t-statistic, PP the Phillips-Perron Z_t statistic. hl(ADF), hl(AR1) and hl(IR) denotes half-lives calculated by the ADF and AR(1) approximation and the impulse response function, respectively. See section A.2 for details. In the tables, ** indicates that the null hypothesis is rejected at a significance level of 1 percent. * denotes significance at the 5 percent level and (*) significance at the 10 percent level.

Table A.1: Average weights, 1973 to 2000

Country	IW	CW	TW
Argentina	0,0015	_	_
Austria	0,0112	_	_
Australia	0,0044	0,0132	0,009
Bangladesh	0,0004	_	_
Belgium	0,0266	0,0291	0,027
Brazil	0,0080	_	_
Canada	0,0212	0,0064	0,023
Switzerland	0,0161	0,0182	0,013
Chile	0,0008	_	_
China	0,0127	_	_
Colombia	0,0018	_	_
Czech Republic	0,0031	0,0009	_
Germany	0,1506	0,1784	0,161
Denmark	0,0721	0,0764	0,068
Spain	0,0124	0,0117	0,012
Finland	0,0376	0,0416	0,034
France	0,0423	0,0501	0,056
United Kingdom	0,1054	0,1154	0,223
Greece	0,0021	0,0028	_
Hong Kong	0,0062	_	_
Hungary	0,0018	0,0004	_
Indonesia	0,0012	_	_
Ireland	0,0081	0,0071	_
India	0,0021	_	_
Iceland	0,0017	_	0,004
Italy	0,0320	0,0356	0,029
Japan	0,0560	0,0649	0,040
S. Korea	0,0093	0,0110	_
Sri Lanka	0,0002	_	_
Morocco	0,0012	_	_
Malaysia	0,0021	_	_
Netherlands	0,0436	0,0439	0,065
Philippines	0,0008	_	_
Pakistan	0,0009	_	_
Poland	0,0074	0,0020	_
Portugal	0,0079	0,0081	_
Russia	0,0144	_	_
Sweden	0,1775	0,2000	0,160
Singapore	0,0043	0,0043	_
Thailand	0,0018	0,0014	_
Turkey	0,0020	_	_
Taiwan	0,0055	0,0044	_
USA	0,0791	0,0726	0,075
South Africa	0,0024	_	_

Average import weights (IW), competitiveness weights (CW) and trade weights (TW) over the period 1973 to 2000.

Table A.2: REER from 1973, with a trend

Group	s.e.	t-ADF	KPSS	trend	hl(ADF)	hl(AR1)	lags
I40	0,160	-4,163*	0,080	-0,0004	0,63	1,01	1
Scandinavia	0,150	-3,172		-0,0005	1,08	1,67	1
Nordic	0,158	-3,426		-0,0004	0,89	1,45	1
S. Europe	0,159	-3,812*	0,102	0,0055*	0,75	1,11	1
N. Europe	0,145	-4,128*	0,073	0,0012(*)	0,76	1,35	1
Top ten	0,146	-4,530**	0,054	0,0021*	0,64	0,98	1
Old industrial	0,152	-4,254*	0,121(*)	0,0015*	0,67	1,00	1
Western	0,147	-4,245*	0,106	0,0005	0,71	1,04	1
Some to 1950	0,162	-4,328**	0,051	0,0019*	0,57	0,81	1
Africa	0,122	-2,372		-0,0037	2,02	2,87	1
Lat.-America	0,153	-2,290		-0,0008	1,60	1,60	0
Less developed	0,182	-2,814		-0,0083*	0,96	0,96	0
Asia	0,177	-3,088		-0,0050(*)	0,88	0,88	0
	1%[**]	-4,324	0,216				
Critical	5%[*]	-3,581	0,146				
values	10%[(*)]	-3,225	0,119				

Table A.3: REER from full sample, without a trend

Group	s.e.	t-ADF	DF-GLS	KPSS	PP	obs	lags	hl(ADF)
Scandinavia	0,088	-2,106	-2,118*	0,559*	-2,091	50	0	3,38
Nordic	0,082	-1,954	-1,746(*)	0,663*	-1,832	50	0	3,99
Some to 1950	0,046	-1,999	-0,932		-2,002	50	0	7,24
I40(–)	0,056	-1,826	-0,856		-1,842	41	0	6,48
S. Europe	0,085	-1,686	-1,591		-1,791	41	0	4,46
Western(-Ger)	0,059	-1,813	-1,074		-1,849	41	0	6,15
Old ind.(-Ger)	0,068	-2,070	-1,703(*)	0,259	-2,116	41	0	4,55
Top nine(-Ger)	0,066	-1,943	-1,194		-1,977	41	0	5,04
Africa	0,044	-1,706	0,135		-1,715	41	0	8,97
Lat.-America	0,076	-1,989	-1,182		-1,989	41	0	4,22
Less developed	0,044	-1,622	-0,606		-1,620	41	0	9,28
Asia	0,121	-2,754	-2,582*	0,407(*)	-2,754(*)	41	0	1,71
N. Europe	0,100	-3,024*	-2,108*	0,257	-2,137	41	1	1,94
	1%[**]	-3,568	-2,613	0,739	-3,571			
Critical	5%[*]	-2,921	-1,948	0,463	-2,922			
values	10%[(*)]	-2,599	-1,613	0,347	-2,599			

A minus sign indicates that some countries have been dropped from the group because data are not available. See section A.4.

Table A.4: REER full sample, with a trend

Group	s.e.	t-ADF	KPSS	trend	DF-GLS	obs	lags	hl(ADF)
Scandinavia	0,111	-3,493(*)	0,088	-0,0009*	-2,844	50	1	1,42
Nordic	0,113	-3,535*	0,106	-0,0010*	-2,941(*)	50	1	1,36
Some to 1950	0,072	-1,389		-0,0001	-1,414	50	0	6,60
I40(–)	0,097	-1,912		-0,0006	-1,916	41	0	3,38
S. Europe	0,088	-1,774		0,0004	-1,816	41	0	4,07
Western (-Ger)	0,087	-1,516		-0,0002	-1,580	41	0	4,91
Top nine (- Ger)	0,098	-1,833		-0,0004	-1,881	41	0	3,49
Lat.-America	0,115	-2,280		-0,0024	-2,361	41	0	2,28
Less developed	0,141	-2,410		-0,0065*	-2,604	41	0	1,67
Asia	0,133	-3,119	0,172*	-0,0018	-3,201*	41	0	1,29
Old ind. (-Ger)	0,084	-1,544		0,0001	-1,514	41	0	4,99
N. Europe	0,106	-3,118		-0,0003	-2,539	41	1	1,72
Africa	0,097	-2,414		-0,0044*	-1,937	41	1	2,61
	1%[**]	-4,157	0,216		-3,770			
Critical	5%[*]	-3,504	0,146		-3,190			
values	10%[(*)]	-3,182	0,119		-2,890			

Table A.5: Bilateral results from 1973 without a trend

Country	ADF-coef.	s.e.	t-ADF	lags	KPSS	DF-GLS	PP
Argentina	0,444	0,171	-3,241*	0	0,235	-3,005**	-3,003*
Austria	0,861	0,091	-1,519	2	0,545*	-0,911	-1,651
Australia	0,825	0,116	-1,513	0	0,577*	-1,487	-1,513
Bangladesh	0,882	0,092	-1,279	0	0,626*	-1,211	-0,840
Belgium	0,715	0,099	-2,878(*)	1	0,149	-2,000*	-1,941
Brazil	0,564	0,162	-2,701(*)	0	0,138	-1,699	-2,759(*)
Canada	0,749	0,101	-2,496	0	0,533*	-1,551	-2,543
Switzerland	0,763	0,097	-2,436	0	0,558*	-1,533	-2,479
Chile	0,755	0,111	-2,205	0	0,375(*)	-1,687	-2,251
China	0,896	0,059	-1,759	0	0,582*	-0,967	-1,808
Colombia	0,877	0,090	-1,361	1	0,238	-1,922(*)	-1,772
Germany	0,575	0,135	-3,159*	1	0,180	-2,103*	-2,120
Denmark	0,755	0,117	-2,088	0	0,454(*)	-1,036	-2,084
Spain	0,838	0,095	-1,703	0	0,520*	-1,345	-1,619
Finland	0,764	0,128	-1,833	2	0,298	-1,487	-2,049
France	0,599	0,157	-2,559	0	0,111	-2,494*	-2,666(*)
UK	0,753	0,104	-2,368	1	0,474*	-1,602	-1,027
Greece	0,879	0,099	-1,226	0	0,445(*)	-1,211	-1,277
Hong Kong	0,940	0,105	-0,577	0	0,314	-0,716	-0,577
Hungary	0,849	0,096	-1,567	1	0,301	-1,038	-1,352
Indonesia	0,886	0,097	-1,174	0	0,612*	-1,114	-1,034
Ireland	0,730	0,100	-2,711	1	0,521*	-1,739	-1,279
India	0,948	0,040	-1,306	1	0,596*	-0,769	-1,171
Iceland	0,362	0,189	-3,373*	0	0,182	-3,286**	-3,281*
Italy	0,878	0,097	-1,262	0	0,502*	-1,297	-1,430
Japan	0,905	0,094	-1,011	0	0,632*	-0,757	-0,536
S. Korea	0,748	0,130	-1,947	0	0,502*	-1,582	-1,556
Sri Lanka	0,800	0,078	-2,556	0	0,442(*)	-1,395	-2,568
Morocco	0,933	0,038	-1,742	0	0,614*	-0,874	-1,686
Malaysia	0,821	0,103	-1,739	0	0,523*	-1,196	-1,812
Netherlands	0,572	0,138	-3,101*	1	0,160	-2,137*	-2,226
Philippines	0,734	0,134	-1,985	0	0,184	-2,008*	-2,110
Pakistan	0,936	0,075	-0,849	0	0,492*	-0,034	-0,966
Portugal	0,902	0,062	-1,575	1	0,352(*)	-1,181	-0,715
Sweden	0,656	0,136	-2,521	0	0,338	-1,991*	-2,351
Singapore	0,678	0,125	-2,582	1	0,191	-1,708	-1,701
Thailand	0,766	0,121	-1,939	0	0,391(*)	-1,685	-1,939
Turkey	0,814	0,109	-1,704	0	0,502*	-1,179	-1,590
USA	0,675	0,111	-2,936(*)	1	0,111	-1,749	-2,564
South Africa	0,746	0,128	-1,989	0	0,216	-1,839	-1,989
Critical values		1%[**]	-3,689		0,739	-2,650	-3,689
		5%[*]	-2,972		0,463	-1,953	-2,972
		10%[(*)]	-2,625		0,347	-1,610	-2,625

Table A.6: Bilateral half-life after 1973 without trend, and data grades

Country	Grade	hl(ADF)	hl(AR1)	Hl(IR)	abs(IR-AR1)	abs(IR-ADF)	lags
Argentina	B	0,85	0,85	0,85	0	0	0
Austria	A	4,64	3,88	2,7	1,18	1,94	2
Australia	A	3,60	3,60	3,60	0	0	0
Bangladesh	C	5,53	5,53	5,53	0	0	0
Belgium	A	2,07	3,38	3,3	0,08	1,23	1
Brazil	C	1,21	1,21	1,21	0	0	0
Canada	A	2,40	2,40	2,40	0	0	0
Switzerland	A	2,56	2,56	2,56	0	0	0
Chile	B	2,47	2,47	2,47	0	0	0
China	C	6,34	6,34	6,34	0	0	0
Colombia	C	3,75	5,30	4,8	0,50	1,05	1
Germany	B	1,25	2,03	2,3	0,27	1,05	1
Denmark	A	2,47	2,47	2,47	0	0	0
Spain	B	3,93	3,93	3,93	0	0	0
Finland	A	2,58	2,11	2,0	0,11	0,58	2
France	A	1,35	1,35	1,35	0	0	0
UK	A	2,45	5,57	3,8	1,77	1,35	1
Greece	B	5,36	5,36	5,36	0	0	0
Hong Kong	A	11,12	11,12	11,12	0	0	0
Hungary	C	4,24	7,04	5,1	1,94	0,86	1
Indonesia	C	5,73	5,73	5,73	0	0	0
Ireland	A	2,20	4,46	3,6	0,86	1,40	1
India	C	13,02	14,03	14,0	0,03	0,98	1
Iceland	B	0,68	0,68	0,68	0	0	0
Italy	A	5,34	5,34	5,34	0	0	0
Japan	A	6,93	6,93	6,93	0	0	0
S. Korea	B	2,38	2,38	2,38	0	0	0
Sri Lanka	C	3,11	3,11	3,11	0	0	0
Morocco	C	10,07	10,07	10,07	0	0	0
Malaysia	C	3,51	3,51	3,51	0	0	0
Netherlands	A	1,24	1,99	2,3	0,31	1,06	1
Philippines	C	2,24	2,24	2,24	0	0	0
Pakistan	C	10,52	10,52	10,52	0	0	0
Portugal	B	6,73	13,44	8,6	4,85	1,87	1
Sweden	A	1,65	1,65	1,65	0	0	0
Singapore	B	1,79	2,82	2,8	0,02	1,03	1
Thailand	C	2,60	2,60	2,60	0	0	0
Turkey	C	3,36	3,36	3,36	0	0	0
USA	A	1,77	2,03	3,70	1,67	1,93	1
South Africa	C	2,36	2,36	2,36	0	0	0
	av.	3,93	4,44	4,22	0,34	0,41	
	med.	2,59	3,37	3,33	0,00	0,00	
	max	13,02	14,03	14,00	4,84	1,94	
	min.	0,68	0,68	0,68	0,00	0,00	

Table A.7: Bilateral results from 1973 with a trend

Country	s.e.	t-ADF	KPSS	trend	lags	obs	hl(ADF)	hl(AR1)
Argentina	0,184	-3,219		0,0040	0	28	0,77	0,77
Austria	0,172	-3,619*	0,070	0,0055*	1	28	0,71	1,50
Australia	0,170	-3,023		-0,0075*	0	28	0,96	0,96
Bangladesh	0,196	-4,055*	0,192*	-0,0242*	1	28	0,44	0,88
Belgium	0,103	-2,884		-0,0005	1	28	1,97	3,31
Brazil	0,165	-2,646		0,0002	0	28	1,21	1,21
Canada	0,151	-2,556		-0,0033	0	28	1,42	1,42
Switzerland	0,175	-2,718		0,0042	0	28	1,08	1,08
Chile	0,143	-2,068		-0,0028	0	28	1,98	1,98
China	0,145	-1,254		-0,0044	0	28	3,45	3,45
Colombia	0,091	-2,396		-0,0029	1	28	2,83	4,62
Germany	0,140	-3,184		0,0007	1	28	1,17	1,92
Denmark	0,169	-3,318(*)		0,0029*	1	28	0,84	1,35
Spain	0,155	-2,097		0,0032	0	28	1,76	1,76
Finland	0,169	-1,349		-0,0001	2	28	2,68	2,06
France	0,160	-2,517		0,0002	0	28	1,35	1,35
UK	0,113	-3,940*		0,0046*	1	28	1,17	1,94
Greece	0,134	-2,086		0,0026	0	28	2,11	2,11
Hong Kong	0,113	-1,305		0,0032	0	28	4,34	4,34
Hungary	0,104	-1,766		0,0030	0	28	3,42	3,42
Indonesia	0,173	-3,352(*)		-0,0252*	0	28	0,80	0,80
Ireland	0,100	-4,770**	0,138(*)	0,0037*	1	28	1,06	1,52
India	0,102	-1,952		-0,0066	1	28	3,13	5,11
Iceland	0,205	-3,304(*)		-0,0007	0	28	0,61	0,61
Italy	0,134	-2,453		0,0039*	0	28	1,74	1,74
Japan	0,196	-3,786*	0,163*	0,0174*	1	28	0,51	0,89
S. Korea	0,173	-2,896		0,0072*	0	28	1,00	1,00
Sri Lanka	0,116	-1,807		-0,0005	0	28	2,96	2,96
Morocco	0,108	-1,700		-0,0035	1	28	3,43	6,02
Malaysia	0,165	-3,138		-0,0090*	1	28	0,95	1,56
Netherlands	0,141	-3,177		-0,0008	1	28	1,16	1,84
Philippines	0,141	-2,077		-0,0016	0	28	1,99	1,99
Pakistan	0,105	-2,903		-0,0088*	1	28	1,90	3,10
Portugal	0,068	-2,139		0,0018	1	28	4,44	5,12
Sweden	0,163	-3,585*	0,118	-0,0020	1	28	0,79	1,26
Singapore	0,125	-2,812		0,0022	1	28	1,60	2,38
Thailand	0,149	-2,245		-0,0025	0	28	1,71	1,71
Turkey	0,174	-2,426		-0,0081	0	28	1,26	1,26
USA	0,123	-2,612		0,0001	1	28	1,78	2,37
South Africa	0,140	-2,841		-0,0028	1	28	1,36	2,01
	1%[**]	-4,324	0,216	Average			1,75	2,17
Critical values	5%[*]	-3,581	0,146	Median			1,39	1,80
	10%[(*)]	-3,225	0,119	Maximum			4,44	6,02
				Minimum			0,44	0,61

Table A.8: Bilateral results from 1950, without a trend

Country	s.e.	t-ADF	KPSS	obs.	lags	hl(ADF)	hl(AR1)
Argentina	0,074	-2,636(*)	0,724*	51	0	3,21	3,21
Austria	0,070	-1,914		51	0	4,84	4,84
Australia	0,081	-0,226		51	4	37,54	7,32
Bangladesh	0,068	-0,544		42	3	18,26	8,34
Belgium	0,054	-2,433		51	1	4,92	5,21
Brazil	0,043	-3,106*	0,684*	51	2	4,84	6,81
Canada	0,034	-0,985		51	0	20,40	20,40
Switzerland	0,045	-0,874		51	0	17,38	17,37
Chile	0,042	-1,404		50	0	11,42	9,73
China	0,027	-1,164		49	4	22,12	47,89
Colombia	0,031	-1,666		51	1	13,10	11,42
Germany	0,132	-3,224 *	0,158	31	1	1,25	1,99
Denmark	0,059	-0,937		51	2	12,10	9,90
Spain	0,072	-2,544		51	1	3,41	4,30
Finland	0,061	-1,901		51	0	5,59	5,59
France	0,046	-1,382		51	0	10,51	10,51
UK	0,059	-2,203		51	1	5,02	6,68
Greece	0,049	-3,817**	0,568*	50	0	3,38	3,38
Hong Kong	0,069	-1,612		41	0	5,89	5,89
Hungary	0,092	-1,339		31	0	5,28	5,28
Indonesia	0,073	-1,349		41	0	6,71	6,71
Ireland	0,071	-2,300		51	1	3,88	4,63
India	0,018	-0,827		51	1	45,84	35,83
Iceland	0,057	-1,813		51	0	6,38	6,38
Italy	0,069	-2,024		51	0	4,62	4,62
Japan	0,053	-0,502		51	2	25,66	25,49
S. Korea	0,094	-2,741(*)	0,183	49	0	2,32	2,32
Sri Lanka	0,025	-1,330		51	0	20,88	20,88
Morocco	0,015	-1,284		51	0	36,18	36,17
Malaysia	0,045	-1,570		47	2	9,53	9,63
Netherlands	0,064	-1,718		51	0	5,90	5,90
Philippines	0,033	-1,906		51	0	10,55	10,55
Pakistan	0,032	-1,241		51	0	17,35	17,35
Portugal	0,035	-1,771		51	1	10,71	6,52
Sweden	0,074	-1,591		50	0	5,52	5,52
Singapore	0,065	-2,069		41	0	4,81	4,81
Thailand	0,036	-0,822		51	0	23,17	23,17
Turkey	0,045	-1,494		51	0	9,87	9,87
USA	0,032	-1,633		51	1	13,02	12,47
South Africa	0,048	-1,290		51	0	10,89	10,89
	1%[**]	-3,568	0,739	Average		12,11	11,39
Critical	5%[*]	-2,921	0,463	Median		9,53	6,81
values	10%[(*)]	-2,599	0,347	Maximum		45,84	47,89
				Minimum		1,25	1,99

Table A.9: Bilateral results from 1950, with a trend

Country	s.e.	t-ADF	KPSS	trend	lags	obs.	Half-life (ADF)
Argentina	0,111	-3,153		-0,0069	0	51	1,61
Austria	0,073	-3,523*	0,070	0,0014*	2	51	2,34
Australia	0,113	-3,400(*)		-0,0034*	0	51	1,43
Bangladesh	0,126	-3,010		-0,0082*	0	42	1,45
Belgium	0,077	-3,373(*)		-0,0013*	1	51	2,30
Brazil	0,070	-1,674		0,0007	2	51	5,60
Canada	0,107	-3,183(*)		-0,0060*	1	51	1,67
Switzerland	0,081	-3,500(*)	0,140(*)	0,0035*	0	51	2,08
Chile	0,128	-3,308(*)		-0,0120*	3	50	1,26
China	0,123	-2,052		-0,0105	4	49	2,38
Colombia	0,068	-2,369		-0,0044	1	51	3,93
Germany	0,138	-3,235(*)		0,0006	1	31	1,18
Denmark	0,113	-4,080 *	0,108	0,0027*	1	51	1,12
Spain	0,074	-2,774		0,0008	1	51	3,02
Finland	0,074	-1,969		-0,0004	0	51	4,37
France	0,080	-1,979		-0,0012	0	51	4,01
UK	0,068	-2,053		-0,0002	1	51	4,58
Greece	0,063	-2,946		0,0000	0	50	3,39
Hong Kong	0,068	-1,116		0,0018*	0	41	8,76
Hungary	0,088	-1,902		0,0028*	0	31	3,77
Indonesia	0,135	-3,226(*)		-0,0132*	0	41	1,21
Ireland	0,080	-2,242		-0,0002	1	51	3,51
India	0,082	-2,655		-0,0081	1	51	2,81
Iceland	0,095	-2,158		-0,0024	0	51	3,04
Italy	0,069	-1,979		0,0004	0	51	4,72
Japan	0,142	-3,373(*)		0,0078*	2	51	1,06
S. Korea	0,096	-2,583		0,0011	0	49	2,43
Sri Lanka	0,079	-1,168		-0,0026	0	51	7,15
Morocco	0,060	-0,656		-0,0008	0	51	17,18
Malaysia	0,127	-3,624*	0,067	-0,0104*	1	47	1,12
Netherlands	0,088	-2,103		0,0007	0	51	3,40
Philippines	0,073	-1,763		-0,0027	0	51	5,01
Pakistan	0,097	-3,121		-0,0096*	1	51	1,92
Portugal	0,044	-1,238		0,0002	1	51	12,44
Sweden	0,111	-3,582*	0,118	-0,0017*	1	50	1,36
Singapore	0,082	-1,469		0,0004	0	41	5,42
Thailand	0,089	-2,413		-0,0044*	0	51	2,87
Turkey	0,106	-2,544		-0,0077*	0	51	2,21
USA	0,075	-1,724		-0,0018	1	51	5,01
South Africa	0,119	-3,862*		-0,0064*	3	51	1,12
	1%[**]	-4,157	0,216	Average			3,63
Critical	5%[*]	-3,504	0,146	Median			2,84
values	10%[(*)]	-3,182	0,119	Maximum			17,18
				Minimum			1,06

A.6 Graphs of real exchange rates

This section contains graphs of all bilateral real exchange rates in I40 over the post Bretton Woods period, and graphs for a number of REERs covering all the years where data were available.

Figure A.1: Bilateral real exchange rates, 1973 - 2000

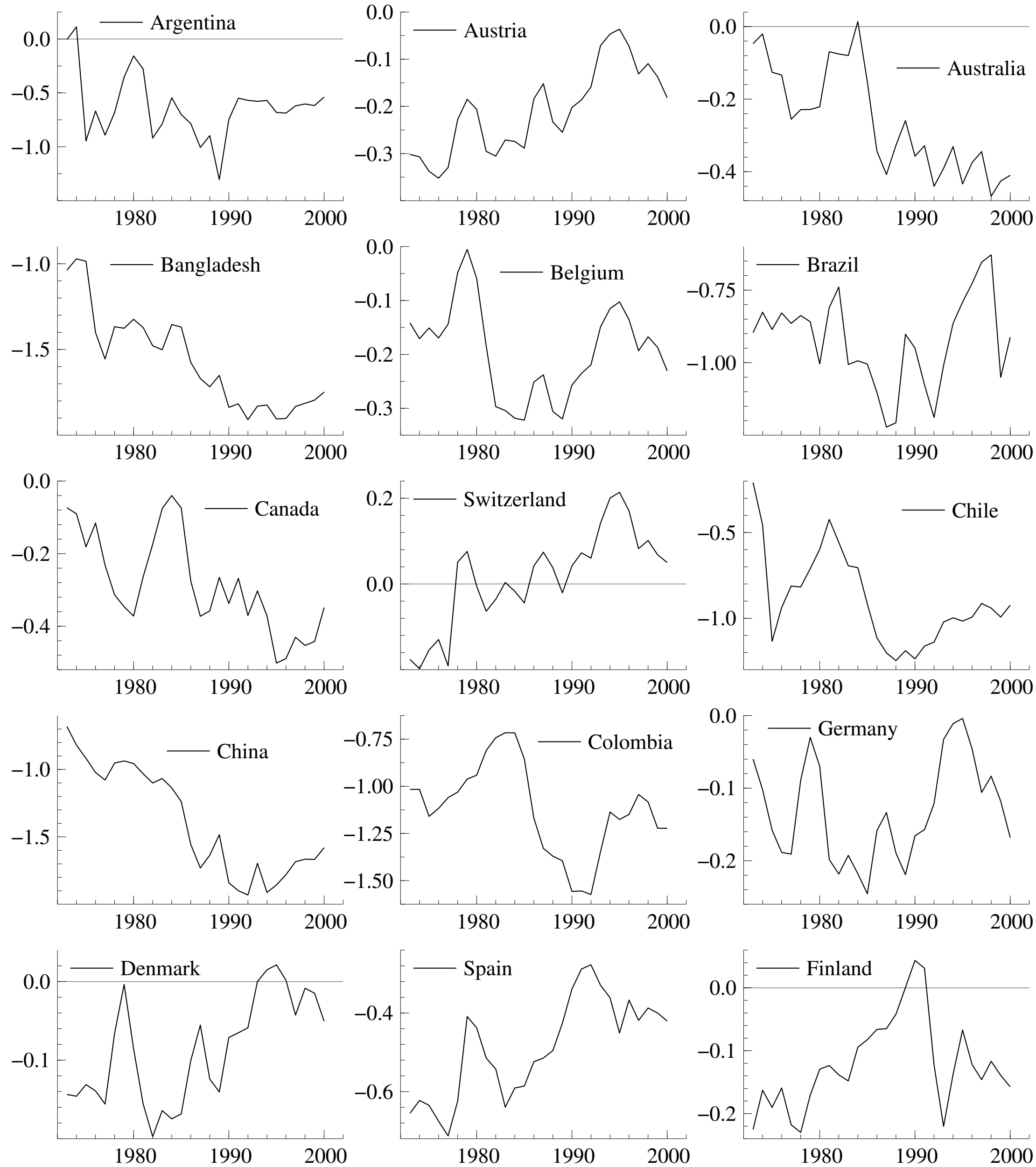


Figure A.2: Bilateral real exchange rates, 1973 - 2000 (continued)

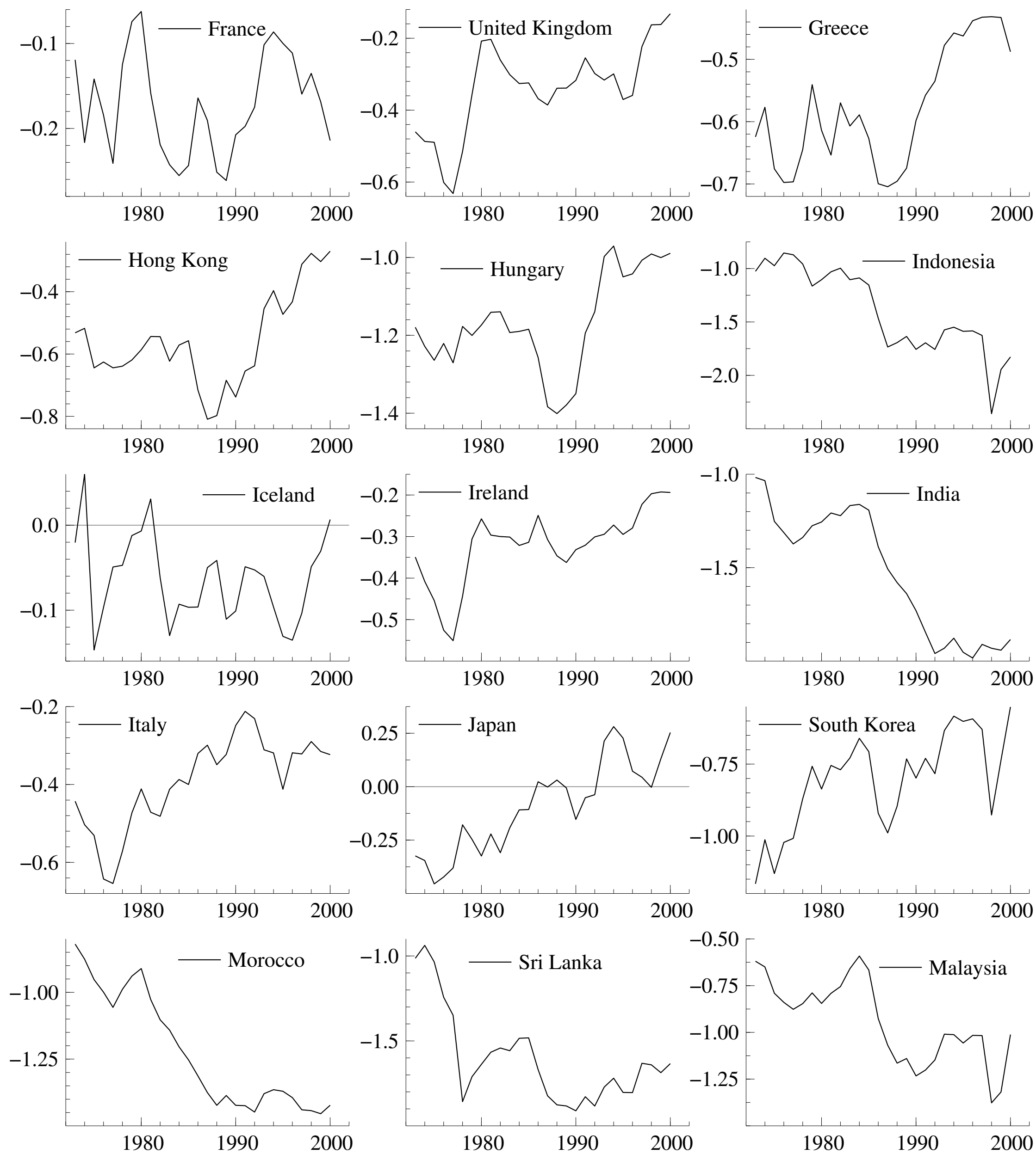


Figure A.3: Bilateral real exchange rates, 1973 - 2000 (cont.)

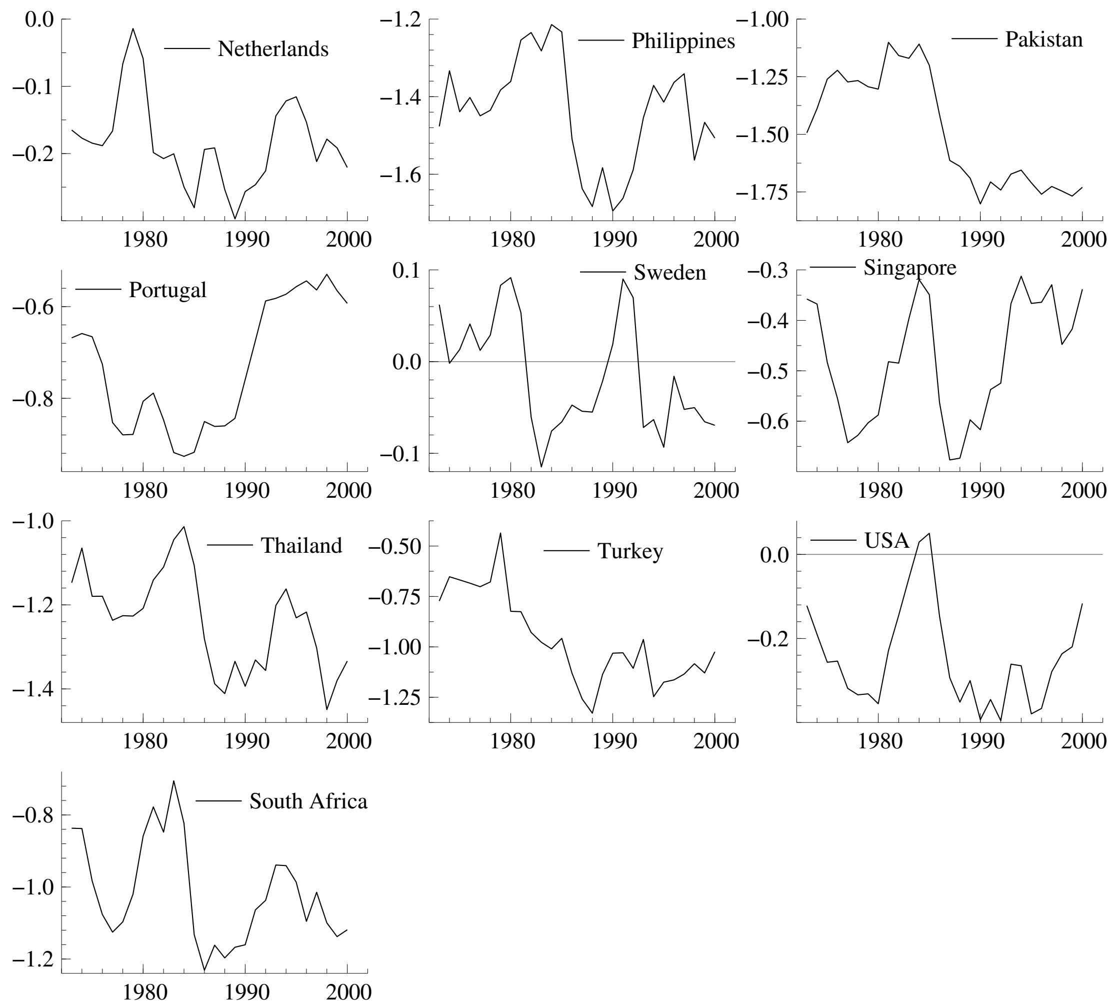
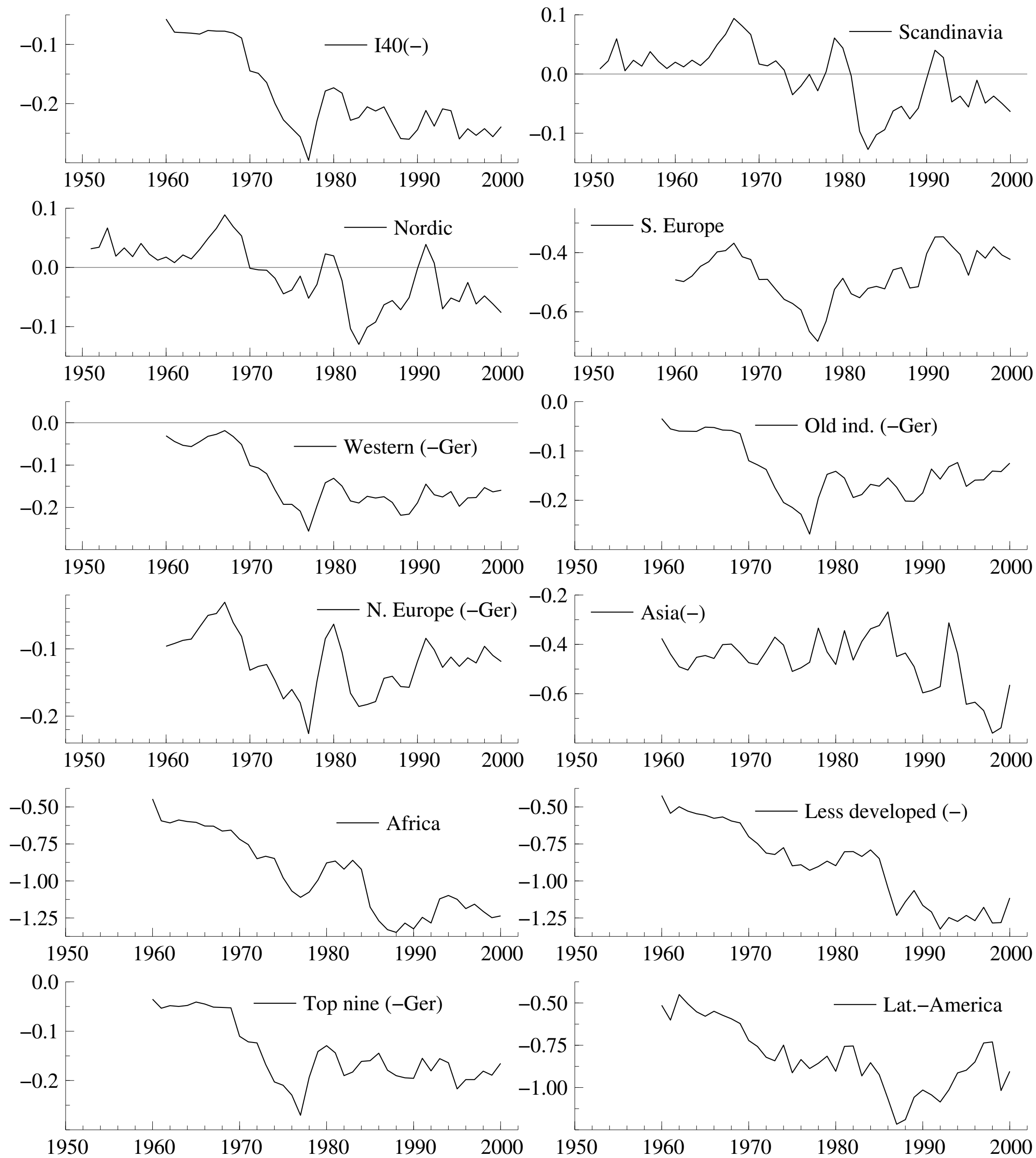


Figure A.4: Real effective exchange rates, full sample



The minus sign in parenthesis indicates that some countries have been dropped from the full sample series as data is not available. See the appendix, section A.4, for details.